Structural and Functional Studies of Eukaryotic and Prokaryotic Bile Salt Transporters from the SLC10, SLC21 and RND Superfamilies

Dissertation
Zur Erlangung der Naturwissenschaftlichen Doktorwürde
(Dr. Sc. Nat.)
vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Universität Zürich
von
Nicole Monroe
aus
Deutschland

Promotionskomitee:
Prof. Dr. Markus G. Grütter (Vorsitz)
Prof. Dr. Bruno Stieger
Prof. Dr. Raimund Dutzler

Zürich, 2010

Summary

Several hurdles have to be overcome in the process of determining the three-dimensional structure of a biomacromolecule. In many cases extensive protein engineering is necessary in order to produce large amounts of material that can be purified to homogeneity. The protein of interest needs to be in its native state, stable and monodisperse. To obtain an atomic model by X-ray diffraction, the generation of well-diffracting crystals is the next bottleneck. Structures of integral membrane proteins are of particular interest because they are the site of action for the majority of drugs. Although amazing accomplishments have been made in the last decade, membrane protein targets remain particularly challenging candidates. This thesis consists of three parts describing advances at various steps of the procedure for eukaryotic and prokaryotic transport proteins from different families.

Health problems related to obeisity, unfavourable blood cholesterol levels and diabetes have become a major concern in the industrialized world. Bile acid homeostasis, being tightly connected to cholesterol levels and the regulation of glucose and lipid metabolism, is thereby of high medical relevance. The sodium taurocholate cotransporting polypeptide NTCP/Ntcp is a key player in the enterohepatic bile salt circulation as it represents the major uptake system for conjugated bile salts into hepatocytes. According to the nomenclature introduced by the Human Genome Organization (HUGO), NTCP/Ntcp is a member of the solute carrier 10 (Slc10) family. The first part of the thesis describes the functional expression of rat Ntcp using baculovirus-infected insect cells. Small-scale purification allowed us to estimate yield and quality of the protein and measures to obtain higher amounts of protein are discussed.

Organic anion transporting polypeptides (OATPs/Oatps), which form the solute carrier 21 (Slc21) superfamily, are also capable of bile acid uptake as part of the enterohepatic bile salt circulation. However, due to their important role in drug disposition, they are mainly studied in the context of drug-transporter or drug-drug interactions. Structural information would greatly advance these investigations. One characteristic structural feature of OATPs/Oatps is a large extracellular loop of about 100 amino acids length. This highly conserved, cysteine-rich loop was predicted to fold into a Kazal-type domain. We therefore designed several constructs for the soluble expression of the proposed domain. While we were able to produce sufficient amounts of pure material, the fragment did not fold into a defined structure. Since the major issue seemed to be the incorrect formation of disulfide bonds, we focused on providing an appropriate redox environment to allow the domain to fold into its native conformation. The biophysical characterization of the extracellular loop of Oatp1a4 is described in chapter 3.

Chapter 4 is concerned with a different kind of bile salt transporter. The multidrug efflux pump AcrB from *Escherichia coli* belongs to the resistance/nodulation/cell division (RND) family and resides in the inner membrane of gram-negative bacteria. As part of a multiprotein complex, it confers resistance to a large number of noxious compounds, including antibiotics, dyes and detergents such as bile salts. Pathogens acquiring resistance to antibiotics are a threat to human health, and the structure and function of this transporter have been investigated extensively.

In our study, AcrB served as a model protein for the use of designed ankyrin repeat proteins (DARPins) to improve the quality of membrane protein crystals. In chapter 4 of this thesis we present a detailed analysis of the interaction of DARPin binders with AcrB. Interestingly, DARPins from the third and fourth round of ribosome display selection were found to bind to their target on a single, overlapping site. In comparing the crystal structures of three AcrB-DARPin complexes, we find that small differences in the binding geometry of seemingly redundant binders significantly affect the crystal packing. This observation advances our understanding of the engineering of crystal contacts. The structures also allow us to learn about the plasticity of a synthetic binding interface. Furthermore, we present a detailed biophysical analysis of the interaction of nine DARPin binders with AcrB. Small variations in the binding surface of the DARPins can cause differences in the affinity and stoichiometry of complex formation and the specificity for binding to AcrB.

Zusammenfassung

Die erfolgreiche Strukturbestimmung eines biologischen Makromoleküls erfolgt in mehreren Schritten, die den Forscher vor Schwierigkeiten stellen können. In vielen Fällen müssen zunächst Eigenschaften des Zielproteins wie Löslichkeit und Stabilität durch Modifikation der Sequenz verbessert werden, um genügend Material produzieren und reinigen zu können. Große Mengen einer möglichst homogenen Proteinlösung, in der das Molekül stabil in seiner nativen Form vorliegt, werden für die Strukturbestimmung benötigt. Eine weitere Schwierigkeit ist die Bildung eines Kristalls mit guten Streueigenschaften für die Röntgenstrukturanalyse, der die Messung der Daten erlaubt, aus denen ein Modell bei ausreichender Auflösung berechnet werden kann.

Die meisten Medikamente wirken durch die Interaktion mit integralen Membranproteinen, die auf der Zelloberfläche zugänglich sind. Daher sind dreidimensionale Strukturen dieser Proteinklasse von besonderem Interesse. Trotz großer Fortschritte in diesem Bereich im letzten Jahrzehnt bleiben Membranproteine eine besondere Herausforderung. Diese Arbeit besteht aus drei Teilen, die Projekte auf verschiedenen Ebenen der Strukturbestimmung beschreiben. Die Zielproteine gehören zu unterschiedlichen Familien eukaryotischer und prokaryotischer Membrantransporter.

In der postindustriellen Gesellschaft der westlichen Welt spielen Gesundheitsrisiken infolge von Übergewicht, hohem Cholesterinspiegel oder Diabetes eine immer größere Rolle. Die Gallensalzhomöostase hat direkten Einfluss auf Cholesterinspiegel und die Regulation des Kohlenhydrat- und Fettstoffwechsels und ist in diesem Zusammenhang von hoher medizinischer Bedeutung. Ein wichtiger Bestandteil des Leber-Darm-Kreislaufs der Gallensäuren ist der Natrium/Taurocholat-Kotransporter NTCP/Ntcp. Dieses Transportprotein gehört zur Familie der natriumabhängigen Gallensalztransporter (Solute carrier 10 family, Slc10) und stellt den wichtigsten Aufnahmeweg für konjugierte Gallensäuren in die Leber dar. Der erste Teil dieser Arbeit befasst sich mit der funktionellen Expression von Ntcp aus der Rattenleber mit Hilfe von baculovirusinfizierten Insektenzellen. Zur Quantifizierung der Proteinmenge wurde auch ein erster Reinigungsversuch im kleinen Maßstab unternommen.

Mitglieder der Organo-Anion-Transporter Proteinfamilie (englisch: organic anion transporting polypeptides, OATPs/Oatps; auch solute carrier 21, Slc21) spielen eine untergeordnete Rolle bei der Aufnahme von Gallensäuren beispielsweise in Leberzellen. Ungleich wichtiger ist jedoch ihr Beitrag zur Entgiftung durch Aufnahme von Fremdstoffen in Leber oder Niere. Da auch die Halbwertszeit von Medikamenten von dem Transport durch OATPs/Oatps abhängt, werden Wechselwirkungen zwischen Substanzen und dem entsprechenden Transportprotein oder auch

der Einfluss eines Medikaments auf den Transport eines anderen Wirkstoffs intensiv untersucht. Diese Studien könnten stark von Proteinstrukturen profitieren, die die Modellierung potentieller Interaktionen ermöglichen würden. Ein charakteristisches Merkmal von OATPs/Oatps ist eine extrazelluläre Domäne zwischen zwei Transmembranhelices, die aus etwa 100 Aminosäuren besteht. Diese Domäne, für die eine Struktur ähnlich der Kazal Serinproteaseinhibitoren postuliert wurde, beinhaltet zehn hochkonservierte Cysteinreste. Kapitel 3 dieser Arbeit beschreibt die Produktion mehrerer Fragmente von Oatp1a4 auf der Grundlage der vorgeschlagenen Struktur mit dem Ziel der Herstellung einer löslichen Domäne. Obwohl genügend Protein hergestellt und zufriedenstellend gereinigt werden konnte, stellte sich heraus, dass das Fragment nicht in der gewünschten Domänenstruktur, sondern in einer ungefalteten Form vorlag. Da die Hauptschwierigkeit in der korrekten Ausbildung der Disulfidbrücken lag, konzentriert sich diese Arbeit auf die Optimierung einer geeigneten Redoxbedingung, die die Faltung des Proteins in die native Konformation ermöglicht. Kapitel 3 beschreibt ausführlich die Analyse des extrazellulären Teils von Oatp1a4 aus der Rattenleber mit Hilfe biophysikalischer Methoden.

Kapitel 4 widmet sich einem andersartigen Gallensalztransporter, nämlich dem Transporter AcrB aus *Escherichia coli*. Dieses Protein gehört zu der Familie von RND Transportern (resistence/nodulation/cell division - Resistenz/Knöllchenbildung/Zellteilung), deren Mitglieder in der inneren Membran gramnegativer Bakterien zu finden sind. In Form eines Proteinkomplexes schützt AcrB die Zelle vor einer Vielzahl schädlicher Substanzen, wie zum Beispiel Antibiotika, Farbstoffe oder Detergentien, indem es diese aus dem Zellinneren ausschleust.

Dieser Teil der Doktorarbeit befasst sich mit der Kristallisation von Membranproteinen unter Zuhilfenahme von Bindungsproteinen, die aus Ankyrinmodulen aufgebaut sind (designed ankyrin repeat proteins, DARPins), zur Verbesserung der Streueigenschaften der Kristalle. AcrB diente uns hierfür als Modellsystem. Die Untersuchung der DARPin Proteine, die durch Evolution im Reagenzglas für die spezifische Bindung an AcrB selektiert wurden, ergab, dass die meisten Binder mit dem selben Teil der Oberfläche des Transporters wechselwirken. Trotz dieser dominanten Bindungstelle macht der Vergleich dreier Kristallstrukturen von AcrB-DARPin Komplexen deutlich, wie unterschiedlich sich ähnliche Bindungspartner auf die Packung und Streueigenschaften des Kristalls auswirken können. Die Strukturen geben ausserdem Einblick in die Variabilität des molekularen Aufbau einer synthetischen Bindungsoberfläche. Für insgesamt neun Komplexe wurde zudem die Interaktion zwischen DARPin und AcrB mit Hilfe von biophysikalischen Methoden untersucht. Auch hier wiesen ähnliche Bindungsmoleküle Unterschiede im Hinblick auf Affinität, Stöchiometrie und Spezifität der Wechselwirkung auf.

Acknowledgements

I would like to thank a number of people who supported me through this work. Prof. Dr. Markus Grütter encouraged me to pursue my research independently. I had the opportunity to attend many practical courses in which I learned a lot. I thank him for his constant support and for providing a stimulating environment that made this work possible.

The projects on eukaryotic bile salt transporters were performed in close collaboration with Prof. Bruno Stieger's group at the Department for Clinical Pharmacology and Toxicology at the University Hospital Zurich. Prof. Stieger was part of my thesis committee and always had time for me when needed. I am grateful for helpful discussions, practical advice in the lab, numerous reference letters and his continuing interest in my work. In the first years of my PhD I worked closely with Anya Hammann-Hänni and I enjoyed the cooperation very much. Stéphanie Häusler introduced me to several kinds of uptake assays for bile acid transporters, and Dr. Christelle Guyot was very helpful with various experiments performed at the University Hospital.

At the Department of Biochemistry, I would like to thank Prof. Raimund Dutzler for being part of my thesis committee. Many members of the Grütter lab deserve special thanks. Dr. Christophe Briand coordinated and assisted in data collection at the synchrotron and furthermore provided an excellent environment for computation. I appreciate his help and patience at all stages of structure determination and refinement. I also want to thank him for teaching me how to perform analytical ultracentrifugation.

I wish to thank all involved in the AcrB project. Gaby Sennhauser initiated the project and passed it on to me along with expertise on purification and crystallization. Discussing the project with her and the other AcrB experts Magdalena Bukowska, Lorenz Brandstätter and Markus Seeger was very helpful. I enjoyed working with Rahel Urech and I would like to thank her for her contribution to the project. Many group members helped me over the years. Daniel Frey first introduced me to the lab. Beat Blattmann and Céline Stutz-Ducommun do an excellent job in providing crystallization service. Sara, Heidi, Magda, Georg, Thomas and Damien often joined me for an early lunch. Nadine deserves credit for answering countless lab questions, bridging long incubation times, introducing me to Zurich's social life and making dinner for me many times. I am also grateful to everybody else who I have become friends with. Your support in and outside the lab was much appreciated. I hope we stay in touch.

I was part of the PhD program in Molecular Life Sciences and I would like to thank Susanna Bachmann for the professional management. Through the program I met fellow researchers that

I am still in touch with as early as during the application phase. The annual retreat was a great opportunity to discuss my research and learn about projects from different fields.

This work would not have been possible without continuing support from my family, Mark, and friends. Thank you so much.

Contents

List of Figures

List of Tables

Abbreviations

ABC	ATP binding cassette
AcrB	Acriflavine resistance protein B
AP	Alkaline phosphatase
Asbt/ASBT	Apical sodium-dependent bile salt transporter
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
BSEP	Bile salt extrusion pump
BSP	Bromosulfophthalein
CD	Circular dichroism
CFTR	Cystic fibrosis transmembrane conductance regulator
DARPin	Designed ankyrin repeat protein
DHEAS	Dehydroepiandrosterone sulfate
DNA	Deoxyribonucleic acid
DPDPE	D-penicillamine-2,5-enkephalin
DTNB	Dithionitrobenzoic acid, Ellman’s reagent
DTT	Dithiothreitol
$E_2 17\beta G$	Estradiol-17-β-glucuronide

EC	Enzyme Commission
ELISA	Enzyme-linked immunosorbent assay
ESI-MS	Electrospray mass spectrometry
FN3 domain	Fibronectin type III domain
FXR	Farnesoid nuclear receptor
GFP	Green fluorescent protein
GPCR	G-protein coupled receptor
GST	Glutathione S-transferase
HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A
HPLC	High performance liquid chromatography
HRP	Horse radish peroxidase
HUGO	Human Genome Organization
IBAT	Ileal bile acid transporter
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LDL	Low density lipoprotein
LTC_4	Leukotriene C4
MBP	Maltose binding protein
MFS	Major facilitator superfamily
mRNA	messenger ribonucleic acid
MRP	Multidrug resistance protein
NMR	Nuclear magnetic resonance
Ntcp/NTCP	Sodium taurocholate cotransporting polypeptide

OAT	Organic anion transporter
Oatp/OATP	Organic anion transporting polypeptide
PAS domain	Per Arnt Sim domain
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEPT	peptide transporter
PFIC	Progressive familial intrahepatic cholestasis
RMSD	Root mean square deviation
RND	Resistance–nodulation–cell division
SCRD	Scavenger receptor cysteine-rich domain
SDS PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SFM	Serum-free medium
SLC	Solute carrier
SOAT	Sodium organic anion transporter
TC(DB)	Transport Classification (Database)
TCA	Trichloroacetic acid
TCDCA	Taurochenodeoxycholate
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TUDCA	Tauroursodeoxycholate

Chapter 1

Introduction

1.1 Classification of membrane proteins

Cells and their organelles are surrounded by a lipid bilayer. Membrane proteins mediate the interaction between the compartments, such as the cytoplasm and the extracellular space. This includes signal transmission and the controlled exchange of polar substances that cannot cross the barrier. Membrane proteins, accounting for roughly one third of the human proteome [1], can be classified in various ways. One possibility is to distinguish between proteins with different secondary structure elements. Beta-barrel proteins such as porins are found in the outer membrane of gram-negative bacteria and in cell organelles. Most integral membrane proteins are alpha-helical. When grouped according to their function, they are roughly divided into receptors, enzymes, channels and transporters. Membrane proteins also play a structural role, act as ligands or mediate vesicle fusion. The following section will focus on transporters, which are the topic of this thesis.

Since the lipid bilayer forms a barrier for polar and charged molecules, their concentration on either side can be substantially different. The physiological extracellular sodium concentration is ten times greater than in the cytoplasm. Conversely, potassium is present at a low concentration in the extracellular fluid, while the concentration in the cell is much higher. This chemical potential, as well as other potentials such as the proton motive force, is used to store and provide energy for biological processes.

Membrane channels and transporters mediate the flux of ions, the uptake of nutrients into the cells, the extrusion of harmful compounds and the generation of concentration gradients across the membrane. The mechanism depends on the direction of transport. Proteins that allow

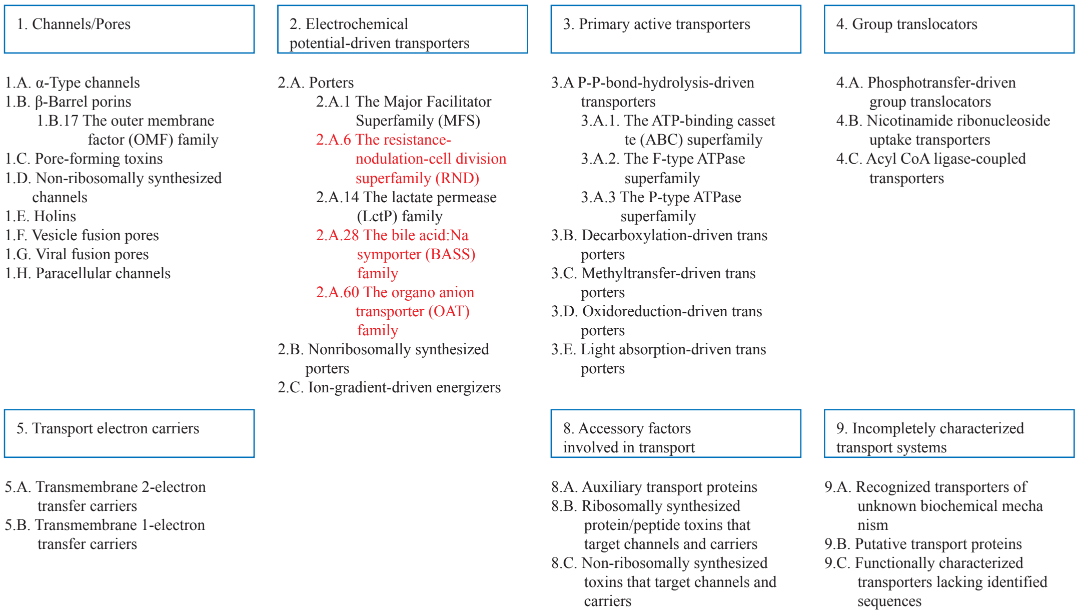


Figure 1.1: Classification of transporters and channels according to the TC system. Transporter families that will be described in this thesis are marked in red.

diffusion along the concentration gradient are named facilitators or passive transporters. Similarly, channels allow passive diffusion of water or a specific ion. In contrast to facilitators, they are characterized by fast rates and a gating mechanism that controls the open probability of the pore. Voltage-gated ion channels open at a certain membrane potential, whereas ligand-gated ion channels conduct in response to ligand binding. Active transport occurs against the concentration gradient of the substrate and therefore requires energy. In the case of primary active transport, the energy is provided by ATP hydrolysis. The $\mathrm{Na}^+\mathrm{K}^+$-ATPase is one of the proteins that maintains the above mentioned ion gradients. A second class of proteins using energy from ATP hydrolysis to move their substrate across the membrane barrier are ATP binding cassette (ABC) transporters. Alternatively, the translocation of the substrate can be coupled to the cotransport or antiport of a second molecule. In this process termed secondary active transport, the energy is provided by the electrochemical gradient of protons or ions.

The human genome project and sequencing of other genomes has provided us with extensive information that allowed the identification of numerous genes. The Gene Nomenclature Committee of the Human Genome Organization (HUGO) grouped human transport proteins into families of solute carriers (SLC) [2,3]. This classification system contains passive transporters and secondary active transporters, but not primary active transporters and ion or water channels. The nomenclature is SLC followed by the number of the family, a letter referring to the subfamily and a number for the individual transporter. Members of the same family share at least 20 % sequence identity. An updated list contains 48 families and 368 transporter genes as of 2010 (http://www.bioparadigms.org/slc/menu.asp). The nomenclature is extended to known homologues of the human proteins. Non-human proteins are designated with lower-case letters. The eukaryotic transporters described in this thesis are members of the sodium-bile salt cotransport superfamily (SLC10/Slc10) and the organic anion transporter family (SLCO/SlcO or SLC21/Slc21).

The Transport Classification (TC) system is an alternative way to group transport proteins and channels (Figure 1.1, similar to the Enzyme Commission (EC) system [4]). A database (Transport Classification Database, TCDB) is maintained online. This nomenclature described in Figure 1.1 is not exclusively based on sequence alignments, but includes experimental information. It applies to all transporters and channels from eukaryotic or prokaryotic origin. Slc10 family members are assigned to the bile acid:Na^+ symporter family, while Slc21 proteins are grouped in the organic anion transporter family together with some members of the Slc22 family. The multidrug efflux pump AcrB described in chapter 4 belongs to the family of resistance-nodulation transporters. All membrane proteins studied in this thesis are electrochemical-potential-driven transporters.

1.2 Advances of Membrane Protein Structural Biology

In 1985, the first high-resolution structure of a membrane protein was solved by X-ray crystallography [5]. The structure of the photosynthetic reaction center from *Rhodopseudomonas viridis* proved that in principle, membrane proteins are amenable to structure determination. In 1988, Johann Deisenhofer, Robert Huber and Hartmut Michel received the Nobel price in Chemistry for this accomplishment. Even though membrane proteins are still underrepresented in the Protein Data Bank (PDB), accounting for 10 % of all structures, the number of membrane protein structures is steadily increasing and has reached 241 unique structures as of 2010 [6].

The progress is due to advances at all stages of the structure determination process. Methods have been developed on the level of target selection, overexpression and crystallization. In the following I will highlight those relevant to this thesis and illustrate them with examples.

1.2.1 Expression strategies

High-resolution structure determination by X-ray crystallography or NMR requires a large quantity of protein in the range of milligrams. The first membrane proteins studied were isolated from tissues where they are abundant in high amounts. Examples are the light-driven proton pump bacteriorhodopsin, which can be purified from the purple membranes of halobacteria. In these membranes, bacteriorhodopsin is present at high concentration in a two-dimensional crystalline arrangement. Rhodopsin is a similar case. The G-protein coupled receptor (GPCR) was isolated from bovine retina. Despite the high relevance of structural information on GPCRs for drug design, rhodopsin was the only receptor that could be isolated in high amounts, making it amenable to structure determination. It is not only due to the progress in protein production, but also the design of more stable variants of the protein and advances in crystallization technology, that high-resolution structures of the human GPCRs β_2-adrenergic receptor [7] and $\mathrm{A}_{2\mathrm{A}}$ adenosine receptor [8] are now available.

Target selection

Highly homologous membrane proteins can behave very differently when produced in a particular expression system. One approach to find a candidate for structure determination is genome pooling. A large number of homologous proteins from different organisms are screened for overexpression. The progress in sequencing of genomes has increased the number of potential targets. If available, bacterial homologs are preferred candidates as they can be more likely produced in a prokaryotic expression system. In particular proteins from thermophilic organisms promise higher stability. Examples for the successful structure determination of a bacterial homolog are the first structures of ABC transporters BtuCD from *Escherichia coli* [9] or Sav1866 from *Staphylococcus aureus* [10].

Expression systems

Advances have been made in the overexpression of recombinant membrane proteins. Bacterial expression systems are preferred because they are cheap and easy to handle. However, since membrane proteins are only stable once they have integrated into the lipid bilayer, overproduction of membrane proteins is often toxic to the expression host and can lead to the formation of intracellular inclusion bodies. *Escherichia coli* strains C41(DE3) and C43(DE3) have been selected for their ability to tolerate membrane protein overproduction [11]. Presumably, these strains can accomodate a larger amount of protein in their membranes.

Eukaryotic expression systems have been used successfully in many cases. Yeast, being the simplest eukaryotic system with well-characterized genetics, has been widely used. The X-ray structure of murine P-glycoprotein was solved recently using protein purified from *Pichia pastoris* for crystallization [12]. The structure is a major accomplishment representing the first structure of a mammalian ABC-transporter. P-glycoprotein, also referred to as multidrug resistance protein 1 (MDR1), is of clinical importance because it limits the bioavailability of drugs. Amongst other substrates, chemotherapeutics are exported out of cells.

Insect cells or mammalian cell lines represent a more complex eukaryotic expression system. One can assume that both the biosynthesis of membrane proteins and the composition of the lipid bilayer are more similar and therefore more suitable for the expression of eukaryotic membrane proteins. Also, these expression systems yield proteins with post-transcriptional modifications such as glycosylation. In particular baculovirus-infected insect cells can yield high amounts of a recombinant protein. Chapter 2 will describe the functional expression of a sodium taurocholate cotransporter from rat liver in baculovirus-infected insect cells.

Engineering

In many cases, engineering of the target is necessary to obtain high expression levels and protein amenable to crystallization. The bottleneck of membrane protein expression is the correct insertion into the membrane. When the N-terminus of the target is located extracellularly, fusion of the terminus to a periplasmic protein such as maltose binding protein (MBP) or thioredoxin has helped integration of the polypeptide chain into the membrane. By this means, GPCR constructs have been identified which can be expressed in *E. coli* even though this protein family is exclusively found in eukaryotes [13].

Like their soluble counterparts, many membrane proteins contain flexible regions that impair stability of the isolated molecule and prevent crystallization. Trimming of the termini or shortening of loops is often necessary. Crystallization of β_2-adrenergic receptor critically depended on engineering of a more rigid construct. In addition to the truncation of the long C-terminus, the dynamic protein was stabilized by replacing a flexible loop with T4 lysozyme [14].

If the protein of interest contains soluble domains, it might be worthwhile to investigate their properties. The nucleotide binding domains of ABC transporters can usually be expressed solubly. X-ray structures of those proteins have provided insight in how the transport is energized.

Similarly, structures of regulatory domains of ion channels have been solved. One example is the PAS domain of hERG [15], a human potassium channel expressed in heart and brain. In chapter 3, the expression of an extracellular domain of organic anion transporter Oatp1a4 is described.

1.2.2 Solubilization and purification

Prior to purification, membrane proteins need to be separated from the lipid bilayer using detergent. The choice of the detergent is critical for the integrity of the membrane protein. While a more powerful detergent might be able to extract more material, it can also interfere with the interactions between the transmembrane segments and thereby disrupt the native fold of the protein. The best choice for solubilization may not be the same detergent the membrane protein tolerates best during purification. For crystallization, yet another detergent may be optimal. This additional parameter which is innate to membrane proteins makes it necessary to screen many conditions for each construct. It is therefore important to optimize the efficiency of the screening procedure, both to find a well-behaving construct and appropriate buffer conditions. A popular approach is to fuse the C-terminus of the candidate to green fluorescent protein (GFP) [16], allowing easy detection of expression, cellular location, solubilization and monodispersity of the unpurified protein by fluorescent size exclusion chromatography [17].

1.2.3 Membrane Protein Crystallization

Lipidic cubic phase

Although crystallization by vapor diffusion has been successful in many cases, some membrane proteins are not stable enough outside of the lipid bilayer. Lipids are necessary for the stability of some membrane proteins. Including phospholipids in the crystallization drop improved or was even required for crystallization in some cases [18,19]. For instance, a study on lactose permease showed how progressive addition of *Escherichia coli* phospholipids during crystallization lead to the formation of different crystal forms with greatly improved diffraction properties [20]. Landau and Rosenbusch demonstrated that membrane proteins can be reconstituted and crystallized in a lipidic bilayer environment [21]. Monoolein spontaneously forms a three-dimensional lipidic cupic phase when mixed with water. Membrane proteins reconstituted into this lipidic bilayer can form crystals by dehydration which induces a transition to a lamellar phase. The first membrane protein structure from lipidic cubic phase crystallization, bacteriorhodopsin, was published in 1997 [22]. The method attracted attention with the crystal structures of the GPCRs

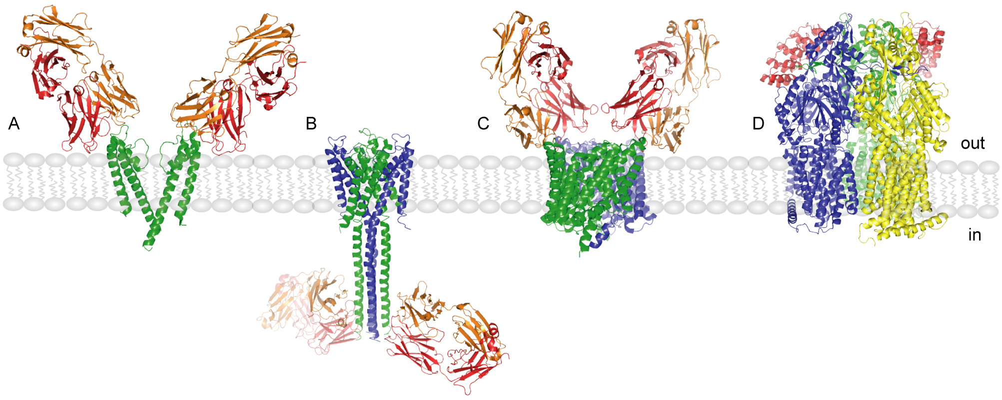


Figure 1.2: Crystal structures of membrane proteins with crystallization chaperones: A) KcsA in complex with Fab fragments. For clarity, only two subunits of the tetramer are shown (PDB ID 1K4C, [24]). B) Full-length KcsA with two Fab fragments (PDB ID 3EFF, [25]). C) ClC (PDB ID 1OTS, [26]). D) AcrB trimer in complex with two DARPin binders (PDB ID 2J8S, [27])

β_2-adrenergic receptor and $\mathrm{A}_{2\mathrm{A}}$ adenosine receptor. Modifications of the lipidic cubic phase method are lipidic sponge phase and lipidic bicelle crystallization. A review of the method providing a list of deposited structures was published by Johanson *et al.* [23].

Crystallization chaperones

Successful crystallization occurs from a supersaturated solution following a nucleation event. Nucleation is more likely to occur in a monodisperse sample. In addition to high purity and the absence of aggregates, conformational homogeneity is desirable. Movement of the protein or high flexibility of loops or termini can prevent crystallization by increasing the entropic cost of lattice formation [28]. Receptors and transporters, however, need to accomodate different conformations during signal transduction or substrate translocation. Accordingly, multiple structures capturing different states of the protein in action allow the understanding of the mode of operation. Ligands or inhibitors are often used to this end. If these are not available or do not lead to successful structure determination for other reasons, artificial binding proteins can be evolved to stabilize a certain conformation. In addition, proteinaceous binding partners increase the hydrophilic surface that is available for crystal contacts. This is important since a large part of the membrane protein surface is shielded by the detergent micelle.

Various scaffolds are available for the evolutionary design of protein binders. In addition to antibody-derived molecules such as Fab or scFv fragments, or the smaller nanobodies, lipocalin-derived anticalins [29], fibronectin type III domain (FN3) and designed ankyrin repeat proteins (DARPins) have been successfully used as binding proteins and crystallization chaperones [30]. DARPins have several advantages over antibody fragments such as exceedingly high stability and high expression yields in a bacterial host [31].

Figure 1.2 shows several crystal structures of membrane proteins in complex with a crystallization chaperone. A high-resolution structure of the transmembrane part of the potassium channel KcsA [24] was obtained with the help of Fab fragments. The Fabs are bound to every chain of the symmetrical tetramer and provide all crystal contacts. The first structure of the full-length channel was solved with the help of Fab fragments targeting the cytoplasmic part of KcsA and stabilizing the closed conformation of the highly dynamic molecule [25]. The crystal structure of *Escherichia coli* ClC was determined at a resolution of 2.5 Å in complex with a Fab fragment, while crystals of the ion channel without binding partner diffracted to 3.5 Å [26]. The multidrug efflux pump AcrB from *Escherichia coli* is the first membrane protein that was crystallized in complex with DARPin binders. The structure provided a higher resolution model of the transporter in an asymmetric conformation [27]. Chapter 4 presents a detailed evaluation of the use of DARPins as crystallization chaperones for AcrB.

Bibliography

[1] Markus Almen, Karl Nordstrom, Robert Fredriksson, and Helgi Schioth. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*, 7:50, 2009.

[2] Matthias A Hediger, Michael F Romero, Ji-Bin Peng, Andreas Rolfs, Hitomi Takanaga, and Elspeth A Bruford. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. *Pflug Arch Eur J Phy*, 447:465–468, 2004.

[3] Avner Schlessinger, Pär Matsson, James E Shima, Ursula Pieper, Sook Wah Yee, Libusha Kelly, Leonard Apeltsin, Robert M Stroud, Thomas E Ferrin, Kathleen M Giacomini, and Andrej Sali. Comparison of human solute carriers. *Protein Science*, 19(3):412–428, 2010.

[4] Milton H Saier, Can V Tran, and Ravi D Barabote. TCDB: the transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Research*, 34:D181–186, 2006.

[5] J. Deisenhofer, O. Epp, K. Miki, R. Huber, and H. Michel. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 åresolution. *Nature*, 318:618–624, 1985.

[6] Stephen H. White. Biophysical dissection of membrane proteins. *Nature*, 459:344–346, 2009.

[7] Vadim Cherezov, Daniel M Rosenbaum, Michael A Hanson, Søren G F Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Peter Kuhn, William I Weis, Brian K Kobilka, and Raymond C Stevens. High-resolution crystal structure of an engineered human beta2-adrenergic g protein-coupled receptor. *Science*, 318:1258–1265, 2007.

[8] Veli-Pekka Jaakola, Mark T Griffith, Michael A Hanson, Vadim Cherezov, Ellen Y T Chien, J Robert Lane, Adriaan P Ijzerman, and Raymond C Stevens. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*, 322:1211–1217, 2008.

[9] Kaspar P Locher, Allen T Lee, and Douglas C Rees. The e. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*, 296(5570):1091–1098, 2002.

[10] Roger J P Dawson and Kaspar P Locher. Structure of a bacterial multidrug ABC transporter. *Nature*, 443:180–185, 2006.

[11] B Miroux and J E Walker. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *JMB*, 260(3):289–298, 1996.

[12] Stephen G Aller, Jodie Yu, Andrew Ward, Yue Weng, Srinivas Chittaboina, Rupeng Zhuo, Patina M Harrell, Yenphuong T Trinh, Qinghai Zhang, Ina L Urbatsch, and Geoffrey Chang. Structure of p-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, 323:1718–1722, 2009.

[13] R Grisshammer, R Duckworth, and R Henderson. Expression of a rat neurotensin receptor in *Escherichia coli*. *Biochem J*, 295 (Pt 2):571–576, 1993.

[14] Daniel M Rosenbaum, Vadim Cherezov, Michael A Hanson, Søren G F Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Xiao-Jie Yao, William I Weis, Raymond C Stevens, and Brian K Kobilka. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science*, 318:1266–1273, 2007.

[15] J H Morais Cabral, A Lee, S L Cohen, B T Chait, M Li, and R Mackinnon. Crystal structure and functional analysis of the HERG potassium channel n terminus: a eukaryotic PAS domain. *Cell*, 95:649–655, 1998.

[16] David Drew, Mirjam Lerch, Edmund Kunji, Dirk-Jan Slotboom, and Jan-Willem de Gier. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods*, 3(4):303–313, 2006.

[17] Toshimitsu Kawate and Eric Gouaux. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure*, 14(4):673–681, 2006.

[18] Huamin Zhang, Genji Kurisu, Janet L Smith, and William A Cramer. A defined protein-detergent-lipid complex for crystallization of integral membrane proteins: The cytochrome b6f complex of oxygenic photosynthesis. *PNAS*, 100(9):5160–5163, 2003.

[19] Marie Jidenko, Rikke C Nielsen, Thomas Lykke-Møller Sørensen, Jesper V Møller, Marc le Maire, Poul Nissen, and Christine Jaxel. Crystallization of a mammalian membrane protein overexpressed in saccharomyces cerevisiae. *PNAS*, 102(33):11687–11691, 2005.

[20] Lan Guan, Irina N Smirnova, Gill Verner, Shushi Nagamori, Shushi Nagamoni, and H Ronald Kaback. Manipulating phospholipids for crystallization of a membrane transport protein. *PNAS*, 103(6):1723–1726, 2006.

[21] E M Landau and J P Rosenbusch. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *PNAS*, 93(25):14532–14535, 1996.

[22] E Pebay-Peyroula, G Rummel, J P Rosenbusch, and E M Landau. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science*, 277(5332):1676–1681, 1997.

[23] Linda C Johansson, Annemarie B Wöhri, Gergely Katona, Sven Engström, and Richard Neutze. Membrane protein crystallization from lipidic phases. *Curr Opin Struct Biol*, 19(4):372–378, 2009.

[24] Yufeng Zhou, Joao H. Morais-Cabral, Amelia Kaufman, and Roderick MacKinnon. Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 åresolution. *Nature*, 414:43–48, 2001.

[25] Serdar Uysal, Valeria Vásquez, Valentina Tereshko, Kaori Esaki, Frederic A Fellouse, Sachdev S Sidhu, Shohei Koide, Eduardo Perozo, and Anthony Kossiakoff. Crystal structure of full-length KcsA in its closed conformation. *PNAS*, 106:6644–6649, 2009.

[26] Raimund Dutzler, Ernest B Campbell, and Roderick MacKinnon. Gating the selectivity filter in ClC chloride channels. *Science*, 300:108–112, 2003.

[27] Gaby Sennhauser, Patrick Amstutz, Christophe Briand, Otso Storchenegger, and Markus G Grütter. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biology*, 5:e7, 2007.

[28] Zygmunt S. Derewenda. Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Cryst D*, 66(5):604–615, 2010.

[29] Arne Skerra. Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS J*, 275(11):2677–2683, 2008.

[30] Shohei Koide. Engineering of recombinant crystallization chaperones. *Current Opinion in Structural Biology*, 19:449–457, 2009.

[31] Gaby Sennhauser and Markus G Grütter. Chaperone-assisted crystallography with DARPins. *Structure*, 16:1443–1453, 2008.

Chapter 2

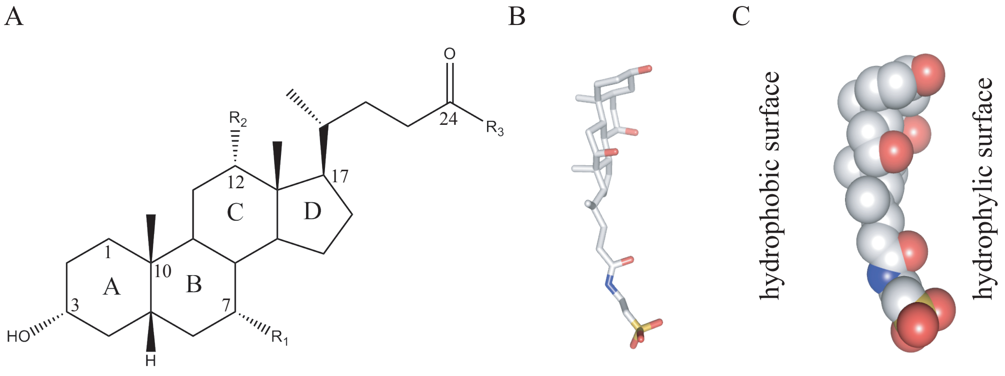
Expression of rat Ntcp in baculovirus-infected insect cells

2.1 Introduction

2.1.1 Bile formation and function

Human bile is formed in two steps: Bile fluid is produced by hepatocytes and modified by cholangiocytes during passage through the long bile ducts. Bile, composed of water, bile acids, phospholipids, cholesterol, proteins and salt, is stored in the gall bladder and released into the duodenum upon stimulus following food-intake. The classical role of bile salts is the emulsification of lipids and lipid-soluble vitamins in the intestine. More recently, additional functions have been discovered. It has been found that bile acids are signalling molecules actively involved in cholesterol metabolism, lipid and carbohydrate homeostasis. Therefore, the bile acid concentration in the intestine and the hepatocyte needs to be tightly controlled.

The synthesis of bile acids is regulated by negative feedback: Binding of bile acids to farnesoid nuclear receptor (FXR) down-regulates the transcription of cholesterol 7α-hydroxylase, the rate-limiting enzyme for bile acid synthesis from cholesterol [1]. Moreover, bile acid activated FXR is a key regulator of genes crucial for hepatic fatty acid and triglyceride synthesis [2]. In addition, bile salts have been found to act as endocrine activators on the G-protein coupled receptor TGR5 [2,3]. In response to bile acid binding, this receptor has a tissue-specific effect on energy and glucose homeostasis. In this context, TGR5 is studied as a potential drug target for the treatment of metabolic disorders such as type 2 diabetes and obesity [4].



	R_1	R_2	R_3
cholic acid	OH	OH	OH
chenodeoxycholic acid	OH	H	OH
deoxycholic acid	H	OH	OH
glycocholate	OH	OH	$NHCH_2CO_2^-$
taurocholate	OH	OH	$NHCH_2CH_2SO_3^-$

Figure 2.1: Structural features of bile acids: A) Chemical formula of common bile acids and their conjugates, adapted from [6] B) Stick representation of taurocholate based on PDB ID 2AZZ [7]. C) Taurocholate molecule in sphere representation illustrating the amphiphatic nature of the molecule.

Primary bile acids are synthesized in hepatocytes from cholesterol [5]. In the classical pathway, 17 enzymatic reactions are required to convert the starting product to a conjugated bile acid. The rate-limiting step is the hydroxylation of C_7 of the steroid nucleus by cholesterol 7α-hydroxylase. Bile acids consist of a sterole with two methyl groups at C_{10} and C_{13}. An isopentenyl or isooctanyl side chain is attached at C_{17} carrying the carboxyl group (C_{24} or C_{27} bile acids depending on the position of the functional group). While all bile acids contain the C_3 hydroxy group of cholesterol and the C_7 hydroxy or oxo group, they differ in the position and orientation of additional hydroxyl groups. The most common bile acids in human bile are chenodeoxycholic acid, cholic acid and deoxycholic acid (Figure 2.1 A). The first two are primary bile acids, while the latter is formed when cholic acid is dehydroxylated by bacteria in the intestine, and is therefore referred to as a secondary bile acid.

While cholesterol is a planar, poorly water-soluble compound that physiologically resides in eukaryotic membranes, bile acids are bent molecules with an apolar and a hydrophilic side (Figure 2.1B and C, [8]) and greatly improved solubility. This detergent-like property enables them to form micelles and emulsify hydrophobic substances such as lipids or lipid-soluble vitamins in the intestine, but also gives rise to the cytotoxicity of bile salts at high concentration. In their physiological form, bile acids are conjugated to the amino acids taurine or glycine. The ionic charge of the resulting bile salts enhances water solubility and prevents the molecules from passing membranes. Protein-mediated transport is therefore required for bile salts to cross the lipid bilayer.

2.1.2 Enterohepatic circulation of bile salts

The major part of bile acids is reabsorbed in the small intestine and returns to the liver after active transport through enterocytes and hepatocytes. This recycling process termed enterohepatic bile salt circulation is tightly connected with lipid homeostasis and cholesterol metabolism. Various reviews describe the enterohepatic circulation of bile salts (e. g. [8,9]). Figure 2.2 gives an overview over the proteins involved in humans as well as their location and driving force. After their synthesis in hepatocytes, glycine- and taurine-conjugated bile acids reach the canaliculus via ATP-hydrolysis-driven transport by the bile salt extrusion pump (BSEP) residing in the canalicular membrane of hepatocytes. Transport occurs against a concentration gradient since bile salts are strongly enriched in the canaliculus. Bile is modified in the bile duct by electrolyte transport proteins such as the cystic fibrosis transmembrane conductance regulator (CFTR) and the chloride/bicarbonate exchanger (AE2) which are expressed in the apical membrane of cholangiocytes (not shown). Bile is stored in the gall bladder and its release into the duodenum is triggered by intestinal hormones. In the small intestine, bile is passed along via peristalsis. In addition to emulsifying water-insoluble food components, bile salts affect digestion, for example through the regulation of pancreatic secretions [10] or activation of carboxyl ester lipase [11].

In the ileum, about 95 % of bile acids are reabsorbed into enterocytes by the apical sodium-dependent bile salt transporter (ASBT). The heterodimeric transporter OSTα/OSTβ [12] translocates the molecules across the basolateral membrane where they are transported back to the liver *via* portal blood. The main uptake system from sinusoidal blood into hepatocytes is the sodium taurocholate cotransporting polypeptide (NTCP). The sodium gradient driving the transport is maintained by the $\mathrm{Na^+K^+}$ ATPase. Sodium-independent organic anion transporting proteins

(OATPs) are also capable of transporting bile salts across the basolateral membrane of hepatocytes, but mainly in their unconjugated forms.

Disturbances of the enterohepatic bile salt circulation are associated with disease. Blockage of bile flux can be hereditary when it is caused by mutation of one of the hepatocyte ABC transporters as it is the case in progressive familial intrahepatic cholestasis (PFIC, for a recent review see [13]). Drug-induced cholestasis, sepsis-associated cholestasis and intrahepatic cholestasis of pregnancy are caused by inhibition or decreased expression level of these transporters [9]. Accumulation of bile acids in hepatocytes ultimately causes liver injury. Malabsorption of bile acids from the intestine leads to cologenic diarrhea. High levels of bile acids are also associated with intestinal inflammation and colon cancer.

Although NTCP/Ntpc and ASBT/Asbt have not been related to disease, their impact on bile salt homeostasis makes them pharmacologically interesting. In industrialized countries, the most common cause of death is cardiovascular disease. High cholesterol level is a risk factor for artheriosclerosis, in particular at higher LDL concentration, increasing the likelihood of heart attack or stroke. Since bile acids are a catabolic end product, their synthesis is the main means of removing cholesterol from the body. This fact has been exploited by blocking the enterohepatic circulation. Drugs targeting ASBT/Asbt aim to prevent artherosclerosis by lowering cholesterol levels in hepatocytes. Several classes of specific inhibitors, some of which are bile salt analogons, met these goals in animal studies (reviewed by [14]). In addition, there is interest in using SLC10 transporters as a shuttle system for prodrugs ([6,15]). Conjugation to bile acids while maintaining the properties of a good ASBT/Asbt substrate can greatly improve the oral bioavailibility of drugs. This approach has been followed by Kramer and Wess. One example is the design of a liver-specific inhibitor of the key enzyme of cholesterol synthesis, HMG-CoA reductase, by combining a statin with a bile acid [16]. The goal is to minimize side-effects in other tissues during cholesterol-lowering therapy.

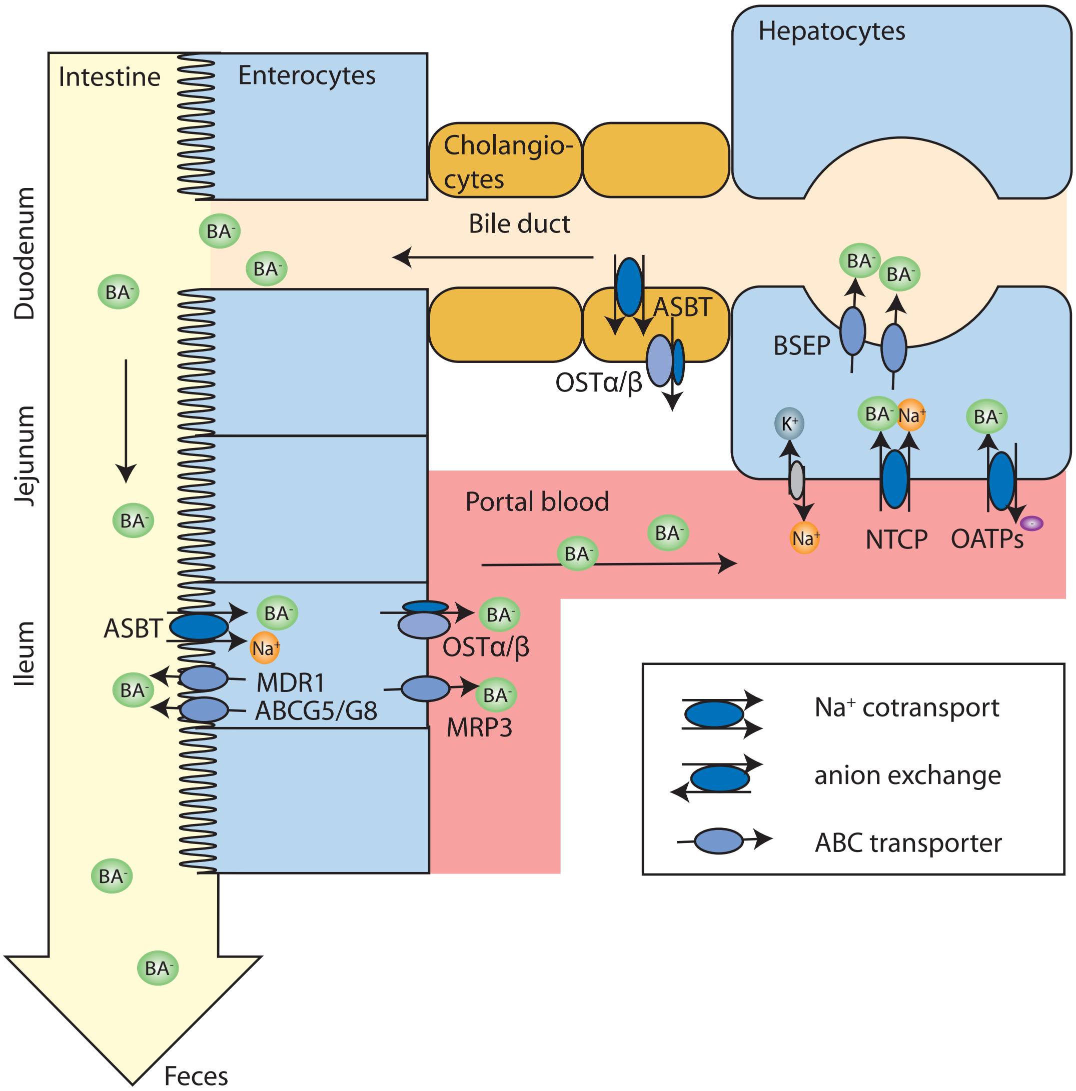


Figure 2.2: Enterohepatic circulation of bile salts: After their synthesis in the liver, bile salts are secreted into the bile canaliculus by ATP-driven transporters such as BSEP, MRP2, MDR3 or ABCG5/G8. Reabsorption in the ileum is driven by ASBT. OATPs and ABC transporters translocate bile acids into the portal blood. From there, uptake into hepatocytes across the basolateral membrane is mediated mainly by NTCP, but also various OATPs.

2.1.3 Sodium-dependent bile acid transporters of the Slc10 family

Sodium-dependent uptake of bile acids which accounts for 80 % of the transport has been studied in basolateral rat liver plasma membrane vesicles [17]. The protein responsible for this activity was later identified by expression cloning [18] in *Xenopus laevis* oocytes. The 51 kDa glycoprotein termed sodium-taurocholate cotransporting polypeptide (NTCP/Ntcp[1]) is the major player of taurocholate uptake across the basolateral membrane of hepatocytes. When the translation of the transporter is blocked by Ntcp-specific antisense oligonucleotides, sodium-dependent taurocholate uptake is decreased by 95 % in rat liver total mRNA injected oocytes ([19]).

The apical sodium bile salt transporter (ASBT/Asbt[1]), also referred to as ileal bile acid transporter (IBAT), is highly homologous to NTCP/Ntcp. Expressed in the brush border membrane of enterocytes, it mediates the reabsorption of bile acids in the ileum. The protein is also found in biliary epithelial cells and renal proximal tubuli.

NTCP/Ntcp and ASBT/Asbt are grouped into the SLC10 superfamily of sodium bile acid transporters that currently holds 29 members, among them seven human proteins. Their transport activity was studied in various eukaryotic cell lines as well as *Xenopus laevis* oocytes. While both NTCP/Ntcp and ASBT/Asbt transport conjugated bile acids with high affinity, the substrate range of NTCP/Ntcp also includes the steroid sulfates estrone-3-sulfate and DHEAS, bromosulfophthalein [20] and the drug-conjugate chloroambucil-taurocholate [21]. Some family members, like the sodium organic anion transporter (SOAT), do not transport bile acids, but other organic anions [22]. The secondary active transporters use the sodium gradient maintained by the Na^+ K^+ ATPase to translocate their substrates in an electrogenic manner, with two sodium ions translocated along with one taurocholate ion [23].

Structural information on SLC10 superfamily members is based on computational sequence analysis. There has been controversy about the number of transmembrane helices. For human ASBT, a seven-transmembrane model was experimentally validated [24]. The N-terminus is located in the extracellular space while the C-terminus is in the cytoplasm. Figure 2.3 shows a 7-helix topology model of rat Ntcp based on a sequence alignment with human ASBT.

Functional analysis of rat Ntcp mutants led to the identification of residues important for transport activity [25]. Two negatively charged residues, D115 and E257, have been found to be crucial for transport. Both are located in loops. This led to the P-loop hypothesis for the transport mechanism. The negatively charged residues are thought to serve as binding sites for

---

[1] All capital letters when referring to the human protein, only the first letter in upper case for its homologs from other organisms.

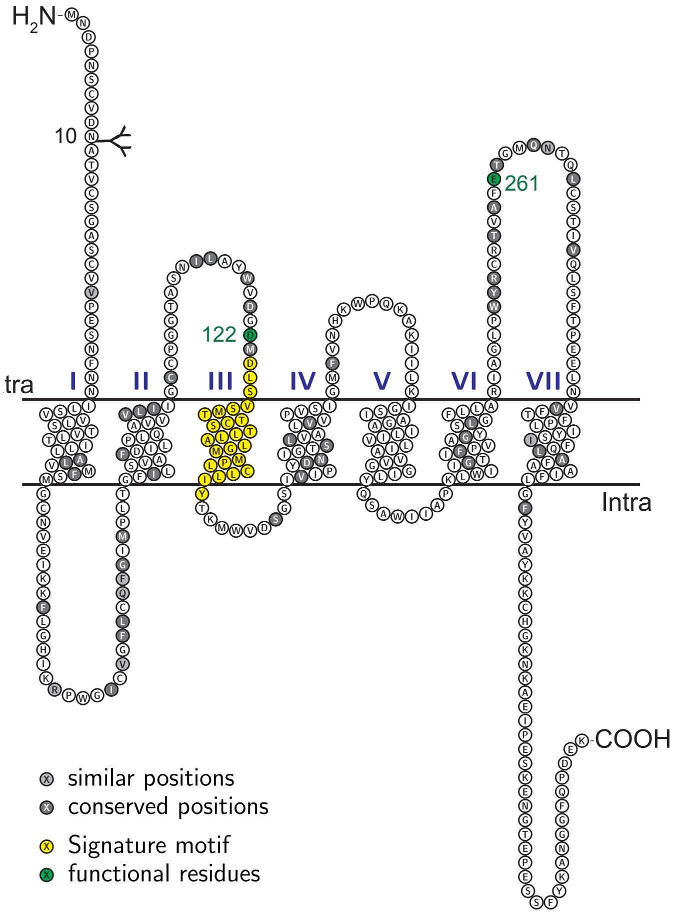


Figure 2.3: Topology model of rat Ntcp (Slc10a1) with seven transmembrane spanning helices. The signature motif for SLC10 family members is colored yellow. A potential glycosylation site is marked in the N-terminal part. Residues D115 and E257 are highlighted in green. Based on a ClustalW alignment of all SLC10 family members, residues are shaded in gray according to their degree of conservation.

the sodium ions. According to the proposed mechanism, the loops containing these residues would insert into the membrane in a P-loop fashion during cotransport of the substrates, similar to the re-entrant loops found in later crystal structures of sodium cotransporters.

A recent publication reported the functional importance of transmembrane helix III in human

ASBT [26]. In particular a proline, which is conserved in all members of the Slc10 family, was identified by the mutational study. Prolines are considered helix breakers because they cannot act as a hydrogen bond donor. A proline in a transmembrane segment would therefore cause a kink or partial unwinding of the helical structure.

2.1.4 Structural features of sodium transporters

Recently, several crystal structures of sodium-dependent bacterial transporters have been determined (Figure 2.4). Although unrelated to the SLC10 superfamily, the structures provide insight into the mechanism of sodium-coupled transport (reviewed by [27]). The glutamate transporter homologue from *Pyrococcus horikoshii* was the first sodium cotransporter with an atomic-resolution structural model (PDB ID 2NWX, [28]). Each chain of the homotrimer contains two re-entrant hairpin loops forming a pseudo-symmetrical structure with an interrupted transmembrane helix. The substrate as well as two sodium binding sites are located in the partly unwound helix. The two hairpin loops prevent the dissociation of glutamate from the transporter. A conformational change would be required to release the substrate.

LeuT (PDB ID 2A65, [29]), vSGLT [30] and Mhp1 [31] are other sodium-coupled transporters with known crystal structures that share a common fold despite a differing number of transmembrane helices. All structures show a pseudosymmetrical motif formed by two bundles of five transmembrane helices (5+5 motif) with a central substrate and ion binding site at the interface of two interrupted transmembrane helices.

The crystal structure of the sodium/H^+ antiporter NhaA (PDB ID 1ZCD, [32]) from *Escherichia coli* shows a 12 transmembrane helix architecture. As observed in the structures of sodium cotransporters, interrupted helices form the binding site for the ions. Disruption of transmembrane helices in ion transporters is therefore a reoccuring observation that might also be present in proteins from the SLC10 family. Discontinuation of the secondary structure leads to a hydrophilic environment caused by exposed main chain atoms, allowing the binding of ions in the otherwise hydrophobic core of the protein.

With the increasing number of membrane protein structures, we understand that the architecture of membrane protein is often much more complex than a simple arrangement of helices perpendicular to the membrane plane. Long helices can cross the membrane at an angle. Re-entrant hairpins, kinks and adjacent helices are possible. In particular the translocation of ions requires a hydrophilic binding site shielded from the lipid environment. Experimental structural data is therefore essential to explain observations from mutational analysis.

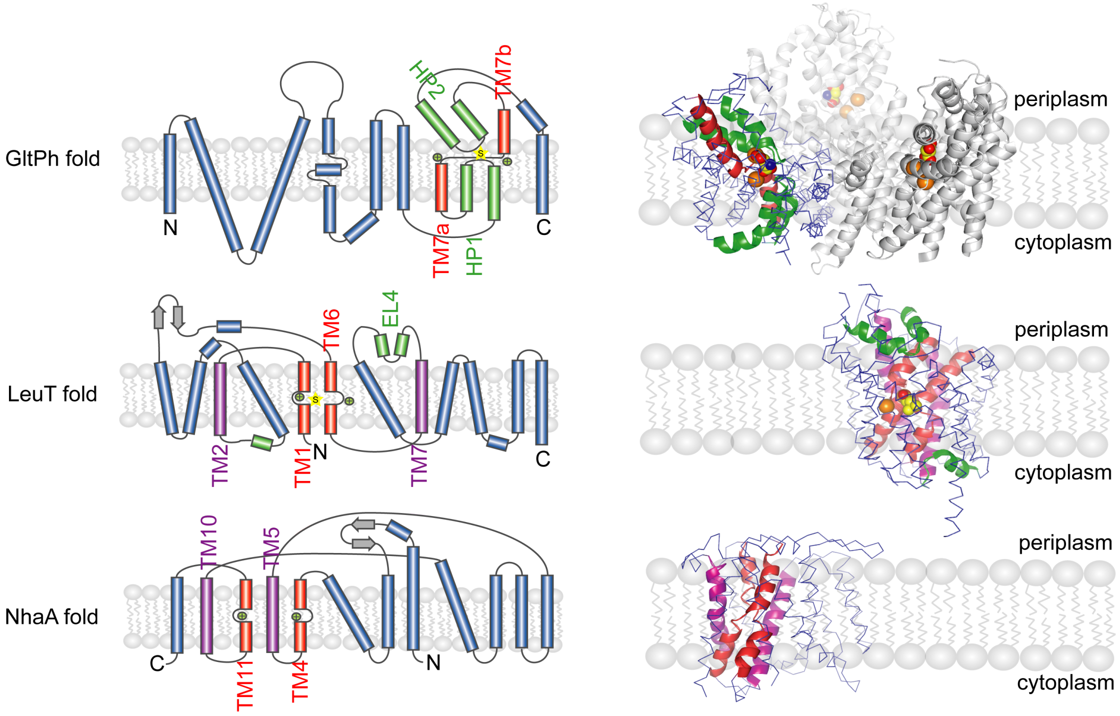


Figure 2.4: Membrane topologies and crystal structures of the sodium-cotransporters GltPh and LeuT as well as the sodium antiporter NhaA. In both the topology and the structure representation, interrupted helices are shown in red, and hairpin loops are colored green. Helices flanking these functional elements are in purple. The rest of the structures is colored in blue. For clarity, one chain of the crystal structures is in ribbon representation, while red, green or purple elements are shown as cartoon. The models for GltPh and LeuT contain bound substrate (yellow) and sodium ions (orange) depicted as spheres.

2.1.5 Aim of the project

Members of the SLC10 superfamily have so far been characterized using transport assays in membrane vesicles. Although membrane preparations can be fractionated, such as during the preparation of basolateral membranes from liver homogenates, the vesicles will contain a multitude of other proteins. In order to study the biochemical and biophysical properties of the isolated transporter, it is desirable to work with pure protein in a solubilized form. Overexpression of functional protein is the first step in providing material for the purification of a protein and ultimately arriving at quantities of homogeneous protein for structural studies.

The aim of this project is the functional overexpression of Ntcp fused to a purification tag using baculovirus-infected insect cells. We evaluated the quantity and quality of the protein after

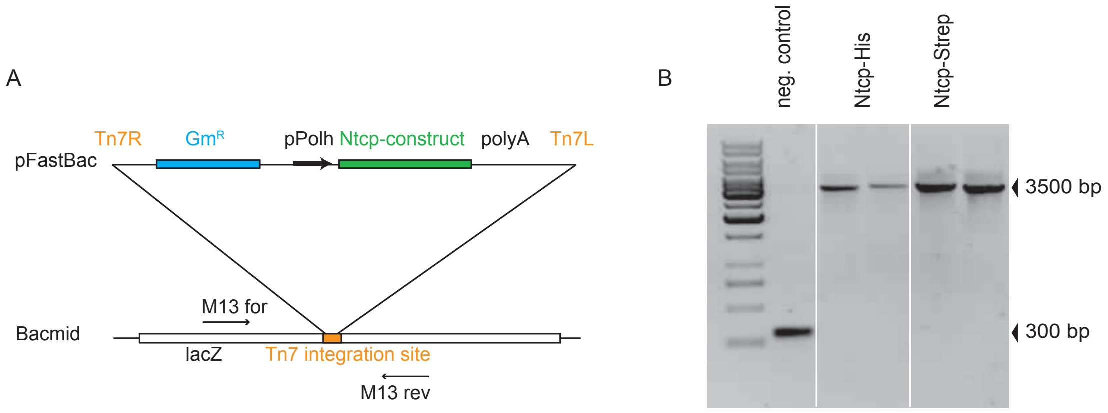


Figure 2.5: Generation of a recombinant bacmid by Tn7 transposition: (A) Schematic of the integration of a gentamycin resistance gene (GmR) along with the Ntcp construct under control of a polyhedrin promoter (pPolh). When integration takes place, a lacZ gene is interrupted. The integration site is flanked by M13 primer binding sites. (B) Analytical PCR using M13 primers. The negative control results in a 300 bp product. Successful transposition with pFastBac-Ntcp-His or pFastBac-Ntcp-Strep yields a PCR product of 3500 bp.

solubilization and purification. Based on the presented experiments, we suggest a strategy for the optimization of the expression and purification of Ntcp for structural studies.

2.2 Results

2.2.1 Generation of recombinant baculovirus stocks

Recombinant bacmids for the expression of full length Ntcp from *Rattus norvegicus* (Slc10a1) were generated by Tn7 transposition in *Escherichia coli* DH10Bac using pFastBac containing the gene modified with a C-terminal *His*- or *Strep*-tag, respectively. Successful transposition leads to the disruption of a *lacZ* gene (Figure 2.5 A), allowing for blue-white screening of transformants. Since the Tn7 site on the bacmid is flanked by M13 primer binding sites, the integration of the gene was confirmed by PCR. Bacmids isolated from white colonies were used as template, and the bacmid from a blue colony served as a negative control. The resulting product is 3500 bp long, confirming the integration of the 1200 bp Ntcp gene along with the flanking sequence, while the negative control shows a 300 bp product (Figure 2.5 B).

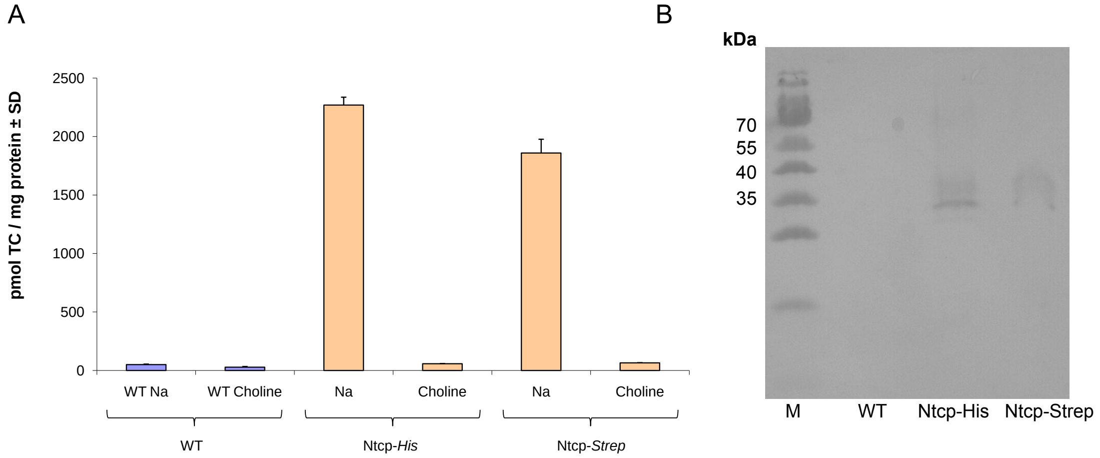


Figure 2.6: (A) Uptake of [^{3}H]-taurocholate by *Sf9* cells three days post-infection. (B) Western blot of Sf9 membranes from cells three days post-infection with WT, Ntcp-His or Ntcp-Strep baculovirus using α-Ntcp antibody.

Table 2.1: Titer of P1 baculovirus stocks

Construct	Virus titer [infectious particles/ml]
Ntcp-His	$1.1 \cdot 10^8$
Ntcp-Strep	$0.8 \cdot 10^8$

Baculovirus stocks were prepared by transfection of insect cells with recombinant bacmid. The titer of the virus stocks after one round of amplification (P1) was around $1 \cdot 10^8$ infectious particles per ml as determined by end-point dilution (Table 2.1). This virus titer is in a good range for infection of insect cells for protein production.

2.2.2 Taurocholate transport by baculovirus-infected *Sf9* cells

Sf9 cells do not contain a sodium-dependent transport system for taurocholate. The presence of functional Ntcp can therefore be demonstrated by uptake of this substrate in a sodium-dependent manner by Ntcp-expressing insect cells 2.6. Three days after infection with Ntcp-containing or wild-type baculovirus, the transport activity of cells in buffer containing sodium was compared to the behaviour in a sodium-free control buffer. The latter contained choline chloride instead of sodium chloride to generate the same ionic strength. Cells infected with wild-type baculovirus did not show a significant uptake of [^{3}H]-taurocholate, whereas in Ntcp-virus-infected cells, the transport activity in sodium-buffer was 30 to 40 times higher than in choline buffer (Figure 2.6 A), confirming the presence of Ntcp in a functional state.

In order to verify the exression of Ntcp, we prepared membrane vesicles from *Sf9* cells infected with Ntcp-baculovirus three days post-infection. A specific antibody for Ntcp detected the protein when analyzed by western blot, while no signal was observed for *Sf9* vesicles from WT-baculovirus infected cells (Figure 2.6 B). The band appears at an apparent molecular weight of 35 kDa, with a larger molecular weight species around 38 kDa. The theoretical molecular mass of rat Ntcp is 39.3 kDa. Due to their hydrophobicity, membrane proteins are known to migrate at a smaller molecular weight than soluble proteins. The presence of two bands suggests that Ntcp partly underwent posttranslational modification such as glycosylation. When detecting Ntcp in basolateral membranes prepared from rat liver homogenates, the apparent molecular weight of the transporter is 51-55 kDa [33]. Hence, the natural protein is subject to much more extensive post-translational modification as compared to Ntcp expressed in insect cells. However, the difference in glycosylation does not seem to prevent taurocholate transport.

2.2.3 Small-scale purification of Ntcp-His

A culture of *Sf9* cells was infected with Ntcp-*His* baculovirus for small-scale purification. We used SMS buffer that has previously been established for transport studies in *Sf9* vesicles. The protein was solubilized from *Sf9* vesicles using foscholine-16. This detergent has been shown to efficiently solubilize the transporter [34]. After affinity chromatography, we analyzed samples at every step of the purification process. Although the protein is not visible on SDS PAGE due to the low amount and poor purity, western blot analysis shows that the protein is enriched in the eluate (Figure 2.7 A). We used analytical size exclusion chromatography to evaluate the monodispersity of the protein (Figure 2.7 B). The fractions were TCA-precipitated and analyzed by western blot using a specific α-Ntcp antibody. Ntcp was detected as a single band eluting between 10 and 15 ml. Only a small part of the protein is found in the void volume of the column which is around 8 ml. We conclude that Ntcp is not aggregated, but exists in a variety of oligomeric states. This inhomogeneity does not seem to arise from post-translational modifications, since only a single band is detected on the western blot. The apparent molecular weight is now between 40 and 50 kDa. This corresponds to the mass expected for Ntcp and may be an effect of TCA precipitation on the migration behaviour of the protein. We estimate that the

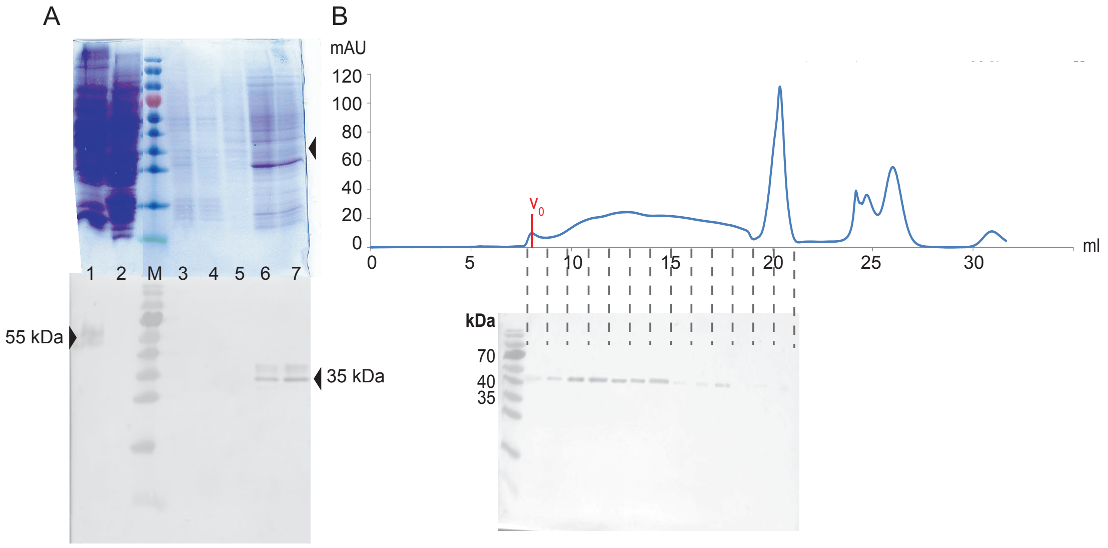


Figure 2.7: A) SDS PAGE and western blot of Ntcp-His at various stages of affinity chromatography: Lane 1) rat liver basolateral membranes, 100 μg total protein; lane 2) WT-baculovirus infected *Sf9* vesicles, 100 μg total protein; lane 3) supernatant after solubilization of *Sf9* cells; lane 4) flow-through of affinity purification using Ni-NTA agarose; lane 5) wash with 50 mM imidazole; lane 6 and 7: fractions of elution with 200 mM imidazole. B) Size exclusion profile of Ntcp-His on a Superdex 200 10/300 GL in SMS, 0.04 % Fos-16 and western blot of fractions.

purification from 100 ml of Sf9 suspension culture yields approximately 100 μg of Ntcp-His.

When repeating the same experiment with Ntcp-*Strep*, we find that the transporter only partly binds to streptavidin matrix during affinity purification and we do not detect protein in the eluate (not shown). Aggregation of this construct might lead to inaccessibility of the purification tag. Since the 9-amino acid Strep tag contains a tryptophan, its hydrophobicity might cause it to interact with the hydrophobic protein or to partition into the detergent micelle.

2.3 Discussion

We have expressed rat Ntcp with a C-terminal purification tag using baculovirus-infected *Sf9* cells. The protein is functional both when fused to a *His*-tag or a *Strep*-tag as shown by a transport assay in whole cells. The identity of the protein was confirmed by western blot analysis using a specific antibody. Functional expression is the first step towards the purification of a protein from a recombinant source.

We were able to solubilize the transporter from insect cell membranes and the protein was enriched during affinity purification via its *His*-tag. The gel-filtration profile showed that the protein eluted in a broad peak, with only a small part in the void-volume. This may be due to multiple oligomeric states of the protein. The native oligomeric state of Slc10 proteins is not known. The yield and purity are poor. However, the amounts of purified protein would be sufficient for functional reconstitution in proteoliposomes. A reconstituted system allows the researcher to control buffer conditions inside and outside of the liposome as well as the membrane composition.

Based on the presented results, we suggest a strategy to improve yield and quality of the produced protein. Since the purification tag does not interfere with functional expression of Ntcp, we propose constructs more suitable for the screening process and for crystallization. A GFP-tagged Ntcp would not only give qualitative information as only correctly folded GFP-fusions show fluorescence, but would also allow for efficient screening of various detergents and buffer conditions monitoring the stability of the protein by fluorescent size exclusion chromatography before purification [35]. The optimization of the detergent is a critical step since not only quantity of extracted protein, but its integrity needs to be considered. The broad elution profile when analyzing Ntcp-His by gel filtration suggests that foscholine-16 disrupts the native state of the transporter and a more mild detergent needs to be used.

Ntcp contains a glycosylation site at position 10 in its N-terminal tail. The corresponding site has been experimentally confirmed for human ASBT, while other potential glycosylation sites were found unmodified [24]. Although various crystal structures of glycosylated proteins have been solved [36], removing glycosylation sites might be necessary to obtain a homogeneous sample for crystallization. On the other hand, glycosylation is often used as quality control in the endoplasmatic reticulum of eukaryotic cells before the protein is incorporated into the plasma membrane. Whether or not unglycosylated Ntcp would be targeted to the plasma membrane can only be found out experimentally. Also here, a GFP-tagged construct would allow to answer this question at an early stage. In addition, Ntcp sequestered intracellularly would not prove functional, as could be easily tested in transport experiments.

The current expression level does not allow for structural studies. Bacterial homologs of sodium taurocholate cotransporters are an alternative. A screen to identify a well-behaving homologous protein that yields high amounts when overexpressed in *Escherichia coli* might lead to faster structure determination. Information on the structure of SLC10 homologs would provide first insight into the architecture and transport mode of SLC10 proteins and allow for the calculation of a more accurate structural model of the eukaryotic transporters.

2.4 Experimental procedures

2.4.1 Maintenance of insect cell cultures

For all experiments, the cell line Sf9 derived from the fall army worm *Spodoptera frugiperda* was used. If not stated otherwise, cells were cultivated in Sf-900 II SFM (serum-free medium, Gibco) supplemented with 100 U/ml penicillin/streptomycin (supplier). All manipulations of insect cell cultures were carried out under sterile conditions in a laminar flow hood.

Thawing of insect cells

To initiate a culture from a frozen stock, the cell suspension was thawed rapidly in a waterbath heated to 37 degrees. After decontamination of the outside of the vial, the cells were immediately transferred into 10 ml of Sf-900 II SFM to dilute the cryoprotectant. After centrifugation at 500 g for 5 min, the cell pellet was carefully resuspended in 10 ml of fresh medium containing 100 U/ml penicillin/streptomycin and transferred into a 10 cm cell culture dish. The cells were incubated in a humidified incubator at 27 degrees.

Passaging

Adherent culture Sf9 cells were grown in monolayers in cell culture dishes (Corning, product number). Per 10 cm dish, $5 \cdot 10^6$ cells were seeded in 10 ml medium. The cells were split at 80 to 100 % confluency to maintain log-phase growth. For this purpose, adherent cells were detached by pipetting and diluted in medium prefilled in fresh cell culture dishes. Typically, 1:3 dilution three times per week were sufficient.

Suspension culture Suspension cultures of *Sf9* cells were grown in glass flasks at 27 degrees and 90 rpm. The cultures were kept at exponential growth by diluting them to $0.8 - 1 \cdot 10^6$ cells per ml every two to three days. Cell densities higher than $3 \cdot 10^6$ cells per ml were avoided. Used flasks were washed with 30 % acetic acid and 70 % ethanol prior to cleaning in the dishwasher with destilled water. The flasks were autoclaved before use.

Freezing of Sf9 cells

To prepare frozen stocks of Sf9 cells, culture with a viability greater than 95 % were used. Cells were harvested at mid log phase growth by centrifugation at 500 g for 5 min and resuspended in cold cryomedium (Sf-900 II SFM supplemented with 10 % DMSO and 100 U/ml penicillin/streptomycin) to a density of $2 \cdot 10^7$ cells/ml. 1 ml aliquots of the cell suspension were prepared in cryo vials and placed in an isopropanol bath for slow cooling to -80 degrees. Frozen Sf9 stocks were ideally stored in liquid nitrogen, alternatively at -80 degrees.

2.4.2 Generation of recombinant baculovirus

Cloning of Ntcp in pFastBac

The full-length rat ntcp gene was amplified by PCR using the primers BamHI-Ntcp and Ntcp-HindIII from the template pSPORT-Ntcp [18]. The resulting product carried the sequence of the Ntcp gene with a BamHI restriction site upstream and a XbaI site, two in-frame stop codons followed by a HindIII restriction site downstream of the open reading frame. The purified PCR product and pFastBac were digested with 5 U BamHI and 5 U HindIII in seperate reactions. After gel-purification of the fragments, the Ntcp gene was ligated into pFastBac.

In order to fuse the Ntcp gene with a C-terminal purification tag, a 6xHis or a Strep-tag with sticky ends corresponding to the restriction enzymes XbaI and HindIII was made by hybridization of the complementary primers Xba-His-for and Xba-His-rev or Strep-for and Strep-rev, respectively. To this end, 1 pmol each of the complementary primers was mixed in 20 mM Tris/HCl pH 7.5, 50 mM NaCl and 2 mM $\mathrm{MgCl_2}$ in a total volume of 10 μl. The hybridization reaction was incubated at 80 degrees for 5 min and then slowly cooled to room temperature. The fragments were ligated in pFastBac-Ntcp digested with XbaI and HindIII. After verifying the sequence using Microsynth standard primers FastBacf and FastBacr, the plasmids could be used for the generation of recombinant bacmid.

Generation of recombinant bacmid

Escherichia coli strain DH10Bac contains a baculovirus shuttle vector (bacmid) that confers kanamycin and gentamycin resistance. A segment of the bacmid encodes the LacZα peptide in which a Tn7 site for bacterial transposition has been inserted. When a gene is integrated in the baculovirus genome at this site, the open reading frame of the LacZα peptide will be disrupted, allowing for blue-white selection of transformants in presence of X-gal. In addition, the integration site is flanked by M13 primer binding sites. The bacteria also harbour a helper plasmid conferring resistance to tetracycline. This plasmid encodes for the transposase catalyzing the Tn7 recombination reaction.

DH10Bac cells were made electrocompetent according to the following procedure: A 100 ml culture of DH10Bac cells in LB medium was grown at 37 degrees and 230 rpm in LB containing kanamycin and tetracyclin. At $OD_{600} = 0.5$, the culture was cooled for 15 min on ice and the cells were pelleted by centrifugation at 4000 rpm for 15 min at 4 degrees. The pellet was resuspended four times in a decreasing volume of ice cold, sterile 10 % glycerol. In the first cycle, 100 ml were used, followed by 50 ml, 2 ml and 200 μl. Competent cells were used immediately or aliquots were snap-frozen in liquid nitrogen and stored at -80 degrees.

20 μl of competent DH10Bac cells were electroporated with 5 ng of the isolated pFastBac plasmid. The cells were immediately resuspended in 4ml 2xYT medium and allowed to recover for 6 h at 37 degrees and 230 rpm. Subsequently, 200 μl of the cell suspension was streaked out on LB plates containing 100 $\mu\text{g/ml}$ ampicillin, 50 $\mu\text{g/ml}$ kanamycin, 10 $\mu\text{g/ml}$ tetracyclin, 7 $\mu\text{g/ml}$ gentamycin, 100 $\mu\text{g/ml}$ X-gal and 1 mM IPTG.

Isolation of the bacmid

Single white colonies were used to inoculate 5 ml LB containing 50 $\mu\text{g/ml}$ kanamycin, 10 $\mu\text{g/ml}$ tetracyclin and 7 $\mu\text{g/ml}$ gentamycin. Blue colonies were used to isolate bacmid for a negative control. The cultures were grown over night at 37 degrees and 230 rpm. In order to isolate the bacmid, the cells were pelleted by centrifugation and resuspended in 300 μl of solution I (15 mM Tris/HCl pH 8.0, 10 mM EDTA, 100 $\mu\text{g/ml}$ RNAse A, filtered 0.2 μm). Alkaline lysis of the cells took place during 5 min after mixing with 300 μl of solution II (0.2 N NaOH, 1 % SDS, filtered 0.2 μm). Then the solution was neutralized by adding of 300 μl of solution III (3 M potassium acetate, pH 5.5, autoclaved). The sample was placed on ice for 10 min. The thick, white precipitate was removed by centrifugation at maximum speed for 10 min. In order to precipitate the bacmid DNA, the supernatant was transferred into a tube containing 800 μl isopropanol. After placing the tube on ice for 10 min, the sample was centrifuged at maximum speed for 15 min. The pellet was washed twice with 70 % ethanol and afterwards allowed to air dry at room temperature for 10 min. The DNA was resuspended in 100 μl 10 mM Tris/HCl pH 8.0 and the concentration determined by measuring the absorption at 260 nm. Purified bacmid DNA was stored at 4 degrees.

Analysis of recombinant bacmid DNA

To verify the integration of the Ntcp gene in the bacmid by Tn7 transposition, analytical PCR was performed making use of the M13 primer binding sites flanking the integration site. 1 μg of bacmid DNA, 1 μl of 10x Thermo Pol buffer, 0.2 μl of 10 mM dNTPs, 0.2 μl of 100 mM

$\mathrm{MgSO_4}$, 2.5 pmol of M13 for, 2.5 pmol of M13 rev and 0.1 μl of Taq polymerase were mixed in a total volume of 10 μl. The reaction was placed in a thermocycler and heated to 93 degrees for 3 min. After 25 cycles of 94 degrees for 45 s, 55 degrees for 45 s and 72 degrees for 5 min, a final extension at 72 degrees for 7 min was performed before cooling to 4 degrees. The products were analysed on a 1 % agarose gel. Blue colonies and negative clones resulted in a PCR product of 300 bp, whereas bacmids successfully transposed with pFastBac produce a band at 2300 bp larger than the size of the insert. In the case of Ntcp, a 3500 bp product confirms the presence of the Ntcp gene on the bacmid.

Transfection

For the transfection of *Sf9* cells with recombinant bacmid, cellfectin (Invitrogen) was used as a transfection agent according to the manufacturer's instructions. Briefly, $9{\cdot}10^5$ cells per well were seeded in sterile 6-well plates with 2ml per well of Baculo Gold medium containing 100 U/ml penicillin/streptomycin. The cells were allowed to attach at 27 degrees for one hour. In the meantime, for each transfection 1 μg of recombinant bacmid DNA diluted in 100 μl Baculo Gold medium was combined with 6 μl of cellfectin diluted in 100 μl medium and incubated for 45 min at room temperature.

Transfections were performed in duplicates with untransfected *Sf9* cells as a control. The cells that should now adhere to the plate were washed once with 2 ml medium. After removing the wash medium, the DNA:lipid complexes were diluted with medium to a total volume of 1 ml and added to the wells containing cells. The plates were incubated at 27 degrees for 5 hours. Then the DNA:lipid complexes were removed by aspiration and 2 ml of Sf-900 serum-free medium containing antibiotics were added. The cells were incubated at 27 degrees until signs of viral infection such as swelling and detachment from the plastic surface became visible. This typically took 3 to 4 days. At this point, the P0 virus stock was recovered. The supernatant was clarified by centrifugation at 500 g and filtered through a 0.2 μm syringe filter. The supernatant constituting the P0 virus stock was stored at 4 degrees protected from light.

Amplification of virus stocks

P0 baculovirus stocks were amplified by infecting a 50 ml suspension culture of Sf9 cells. For each amplification, a culture containing $0.6 \cdot 10^6$ cells per ml was infected with 250 μl of the P0 virus stock. One flask was left uninfected for comparison. The cultures were incubated at 27 degrees and 90 rpm and the number of cells and their viability was monitored. To this end, an aliquot of the cell suspension was mixed with equal volume of Trypan blue. This dye selectively stains dead cells. The cells were counted using a Thomas counting chamber and the fraction of viable cells was determined by dividing the number of cells that were able to exclude Trypan blue by the total number of cells. Virus was harvested when the cells had stopped growing for 24 hours and the viability started to decrease. To recover the P1 virus stock, the culture was centrifuged for 5 min at 500 g and the supernatant was filtered using a 0.2 $\mu\mathrm{m}$ syringe filter. The titer was determined as described in 2.4.2 and the virus stock was stored at 4 or -80 degrees.

Titer determination by end point dilution

In order to determine the titer of a baculovirus stock, a Sf9 cell suspension with $2.5 \cdot 10^4$ cells/ml was prepared and 200 μl were added to each well of a 96-well plate suitable for tissue culture. The cells were allowed to adhere for 1 hour at 27 degrees. 100 μl of a 1 in 1000 dilution of the virus stock in medium was added to the central 6 wells (B1-G1) of the first column of the plate. Wells A1 and H1 served as controls after addition of 100 μl medium. A 12 step 3 times serial dilution of the medium containing the virus stock was prepared by transferring 100 μl from the first to the second column and repeating this step until column 12 was reached. The plates were incubated at 27 degrees in a humidified incubator for 12 days to allow virus particles to infect the cells. After this time, wells that contained any baculovirus particles would not hold viable cells. Thiazolyl Blue Tetrazolium Blue (MTT, Sigma) was used to detect cell viability. The reagent is a yellow substance that is converted to a water-insoluble, dark blue MTT-formazan by mitochondrial dehydrogenases of living cells. 25 μl of a 5 mg/ml MTT solution were added to each well. The plates were wrapped in tinfoil to prevent degradation of the light-sensitive dye and incubated for 3 h at 27 degrees. Subsequently, the supernatant was aspirated and the blue precipitate dissolved in 95 % isopropanol, 50 mM HCl. After 15 min at room temperature, the absorbtion at 550 nm was read in a microplate reader. The first virus dilution where 2 of the 6 wells do not show viral infection was multiplied by 5 to calculate the virus titer. A titer ranging from $1 - 9 \cdot 10^8$ infectious particles per ml is desirable.

2.4.3 Assay to measure uptake into Sf9 cells

To confirm the functional expression of Ntcp in Sf9 cells, we measured sodium-dependent transport activity of Sf9 cells using radioactive taurocholate. Sf9 cells for this experiment were grown in monolayers in Grace's medium supplemented with 10 % FCS and 100 U/ml penicillin/streptomycin. For each sample, three 10 cm Corning dishes with $5 \cdot 10^6$ ml Sf9 cells were infected with the Ntcp or wild type baculovirus stock at a *multiplicity of infection* (MOI) of 5. Three days post-infection, the cells were harvested under non-sterile conditions using a cell scraper. Cells infected with the same virus were pooled and counted. For each sample, six eppendorf tubes were prepared with $1 \cdot 10^6$ cells each. Three of the tubes served to measure the transport in sodium buffer (122 mM NaCl, 10 mM MES, 3 mM $\mathrm{CaCl_2 \cdot 2H_20}$, 5 mM $\mathrm{MgCl_2 \cdot 6H_20}$, 5 mM D-glucose, pH 6.5) and the other three were used as a negative control in choline buffer (122 mM choline chloride, 10 mM MES, 3 mM $\mathrm{CaCl_2 \cdot 2H_20}$, 5 mM $\mathrm{MgCl_2 \cdot 6H_20}$, 5 mM D-glucose, pH 6.5). The cells were washed twice by centrifugation at 1000 g, aspiration of the medium and resuspending in 1 ml Grace's medium supplemented with 100 U/ml penicillin/streptomycin to remove residual FCS. To start the uptake process, the cell pellet was resuspended in 250 μl 10 μM [^{3}H]-taurocholate (2.5 μCi/ml in sodium or choline buffer. After incubating the tube for 15 min at 25 degrees and 800 rpm, the uptake was stopped by addition of 1 ml ice cold choline buffer. The tubes were stored on ice until all samples were processed. To remove remaining extracellular [^{3}H]-taurocholate, the cells were washed twice with 1 ml cold choline buffer by centrifugation for 1 min at 1500 g and 4 degrees, aspiration of the supernatant and resuspending the pellet. After an additional centrifugation step, the cells were lysed in 250 μl of 1 % triton X-100 during 1 h at room temperature. The total protein concentration of each sample was determined in a BCA assay according the the manufacturer's instructions using 25 μl of the lysate. 150 μl of the lysate were mixed with 4 ml Ultima Gold in a scintillation counter vial. To relate the radioactivity count to taurocholate concentration, the [^{3}H]-taurocholate solution in sodium or choline buffer, respectively, was used as a standard. To this end, 50 μl of each solution were added to 100 μl triton X-100 and 4 ml Ultima Gold in a scintillation vial. The standards were prepared in triplicates. A scintillation counter was used to measure the radioactivity from tricium for 5 min per vial. The transport activity was calculated in pmol taurocholate per mg total protein. For each sample, the mean transport activity in sodium buffer was compared to the signal in absence of sodium.

2.4.4 Small-scale preparation of vesicles

Vesicles from Sf9 cells were prepared as a partial purification step to be used as samples for SDS PAGE and western blot analysis, for transport experiments (not described) or for solubilization of the transporter. Cells were harvested by centrifugation at 3000 rpm for 10 min at 4 degrees and resuspended in 5 ml cold TMEP (50 mM Tris/HCl pH 7.0, 50 mM mannitol, 2 mM EGTA, Complete) per $5 \cdot 10^7$ cells. The cells were homogenized using a glass-teflon homogenizer and a drill for 20 strokes at full speed while keeping the samples on ice. After centrifugation at 3000 rpm for 10 min at 4 degrees, the supernatant was transferred to ultracentrifugation tubes. Following centrifugation at 100000 g for 1 h at 4 degrees, the membrane pellet was resuspended in 100 μl SMS buffer (20 mM HEPES/Tris pH 7.5, 250 mM sucrose, 0.2 mM $\mathrm{CaCl_2}$) with a 25 G needle until the solution turned turbid. The total protein concentration was determined using a BCA assay and the sample stored at -80 degrees after snap-freezing in liquid nitrogen.

2.4.5 Protein purification and analysis

Infection

For small-scale protein expression, a 100 ml suspension culture of Sf9 cells with a density of $1 \cdot 10^6$ cells/ml was infected using the respective P1 virus stock at a MOI of 5. The cell growth was monitored by performing a viable cell count every 24 h. The cells were harvested at an early state of infection, typically after 72 h. This state is reached when enlarged cells are observed and the culture stopped growing. It is important to harvest before the infection causes lysis of the culture.

Harvesting and membrane preparation

The cells were harvested by centrifugation at 5000 rpm in a 50 ml Falcon tube. After resuspending the pellet in 20 ml cold TMEP (50 mM Tris/HCl pH 7.0, 50 mM mannitol, 2 mM EGTA, Complete), the cells were lysed using three rounds of French Press at 4 degrees. Cell debris was removed by centrifugation at 5000 rpm for 10 min at 4 degrees and the supernatant was subjected to ultracentrifugation at 45000 rpm (100000 g, Beckman Ti70) for 1 h at 4 degrees. The membrane pellet was resuspended in 500 μl SMS buffer (20 mM HEPES/Tris pH 7.5, 250 mM sucrose, 0.2 mM $\mathrm{CaCl_2}$). The membrane suspension was snap-frozen in liquid nitrogen for storage at -80 degrees or directly used for solubilization.

Solubilization and purification

The membrane suspension was solubilized in SMS supplemented with 1 % foscholine-16 during overhead rotation over night at 4 degrees in a total volume of 30 ml. Unsolubilized material was removed by ultracentrifugation and the supernatant was subjected to affinity chromatography.

Ntcp-*His* was purified by batch-binding to 500 μl Ni-NTA agarose equilibrated with SMS, 10 mM imidazole, 0.04 % Fos-16 for 30 min at 4 degrees. The suspension was poured into a gravity-flow column, washed with 10 column volumes of SMS, 50 mM imidazole, 0.04 % Fos-16 and eluted in SMS, 200 mM imidazole, 0.04 % Fos-16.

The purification of Ntcp-*Strep* was carried out similarly by adding 500 μl of StrepTactin (IBA Biotech, Germany) equilibrated with SMS, 0.04 % Fos-16 to the supernatant after ultracentrifugation. After batch-binding for 30 min at 4 degrees, the suspension was poured into a gravity-flow column and the matrix was washed with 10 column volumes of SMS, 0.04 % Fos-16. 2.5 mM desthiobiotin in SMS, 0.04 % Fos-16 was used to elute the protein from the column.

The fractions containing the most protein as judged by measuring absorbance at 280 nm were analyzed by size exclusion chromatography in SMS, 0.04 % Fos-16 using a Superdex 200 10/30 GL column.

2.4.6 SDS PAGE and Western Blot

Protein samples were analyzed by SDS PAGE using 10-20 % gradient gels. The gels were either stained with Coomassie Brilliant Blue or the protein was transferred to a nitrocellulose membrane using the iBlot Dry Blotting System (Invitrogen). The membranes were blocked in high TBS-T (50 mM Tris/HCl pH 7.0, 500 mM NaCl, 0.2 % Tween20) containing 5 % milk overnight at 4 degrees. After washing twice for 10 min with high TBS-T, the membranes were incubated with the primary antibody (K4 rabbit α-Ntcp, 1:5000 in high TBS-T/0.5 % milk) for 2 h at room temperature. Following an additional washing step, the secondary antibody (goat α-rabbit IgG horse radish peroxidase, Sigma, 1:5000 in high TBS-T/0.5 % milk) was applied for 1 h at room temperature. Subsequently, the membranes were washed with high TBS-T, high TBS (50 mM Tris/HCl pH 7.0, 500 mM NaCl) and destilled water twice for 5 min each. The blots were developed using the ECl kit (GE Healthcare) according to the instructions and chemiluminescence was read using a LAS-3000 Luminescent Image Analyzer (FujiFilm).

Bibliography

[1] John Y L Chiang. Regulation of bile acid synthesis: pathways, nuclear receptors, and mechanisms. *J Hepatol*, 40:539–551, 2004.

[2] Verena Keitel, Ralf Kubitz, and Dieter Häussinger. Endocrine and paracrine role of bile acids. *World J Gastroentero*, 14:5620–5629, 2008.

[3] Yuji Kawamata, Ryo Fujii, Masaki Hosoya, Masataka Harada, Hiromi Yoshida, Masanori Miwa, Shoji Fukusumi, Yugo Habata, Takashi Itoh, Yasushi Shintani, Shuji Hinuma, Yukio Fujisawa, and Masahiko Fujino. A G protein-coupled receptor responsive to bile acids. *JBC*, 278:9435–9440, 2003.

[4] Atul Tiwari and Pranab Maiti. TGR5: an emerging bile acid G-protein-coupled receptor target for the potential treatment of metabolic disorders. *Drug Discov Today*, 14:523–530, 2009.

[5] David W Russell. Fifty years of advances in bile acid synthesis and metabolism. *J Lipid Res*, 50 Suppl:S120–125, 2009.

[6] Elina Sievänen. Exploitation of bile acid transport systems in prodrug design. *Molecules*, 12:1859–1889, 2007.

[7] Ying H Pan and Brian J Bahnson. Structural basis for bile salt inhibition of pancreatic phospholipase a2. *JMB*, 369(2):439–450, 2007.

[8] A F Hofmann and L R Hagey. Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. *Cell Mol Life Sci*, 65:2461–2483, 2008.

[9] Gerd A Kullak-Ublick, Bruno Stieger, and Peter J Meier. Enterohepatic bile salt transporters in normal physiology and liver disease. *Gastroenterology*, 126:322–342, 2004.

[10] I Koop, M Schindler, A Bosshammer, J Scheibner, E Stange, and H Koop. Physiological control of cholecystokinin release and pancreatic enzyme secretion by intraduodenal bile acids. *Gut*, 39:661–667, 1996.

[11] R Jason Kirby, Shuqin Zheng, Patrick Tso, Philip N Howles, and David Y Hui. Bile salt-stimulated carboxyl ester lipase influences lipoprotein assembly and secretion in intestine: a process mediated via ceramide hydrolysis. *JBC*, 277:4104–4109, 2002.

[12] Nazzareno Ballatori, Na Li, Fang Fang, James L Boyer, Whitney V Christian, and Christine L Hammond. OST alpha-OST beta: a key membrane transporter of bile acids and conjugated steroids. *Front Biosci*, 14:2829–2844, 2009.

[13] Anne Davit-Spraul, Emmanuel Gonzales, Christiane Baussan, and Emmanuel Jacquemin. Progressive familial intrahepatic cholestasis. *Orphanet J Rare Dis*, 4:1, 2009.

[14] J. Geyer, T. Wilke, and E. Petzinger. The solute carrier family SLC10: more than a family of bile acid transporters regarding function and phylogenetic relationships. *N-S Arch Pharmacol*, 372:413–431, 2006.

[15] Anand Balakrishnan and James E Polli. Apical sodium dependent bile acid transporter (ASBT, SLC10A2): a potential prodrug target. *Mol Pharm*, 3:223–230, 2006.

[16] W Kramer, G Wess, A Enhsen, K Bock, E Falk, A Hoffmann, G Neckermann, D Gantz, S Schulz, and L Nickau. Bile acid derived HMG-CoA reductase inhibitors. *BBA*, 1227:137–154, 1994.

[17] B Zimmerli, J Valantinas, and P J Meier. Multispecificity of Na+-dependent taurocholate uptake in basolateral (sinusoidal) rat liver plasma membrane vesicles. *J Pharmacol Exp Ther*, 250:301–308, 1989.

[18] B Hagenbuch, B Stieger, M Foguet, H Lübbert, and P J Meier. Functional expression cloning and characterization of the hepatocyte Na+/bile acid cotransport system. *PNAS*, 88:10629–10633, 1991.

[19] B Hagenbuch, B F Scharschmidt, and P J Meier. Effect of antisense oligonucleotides on the expression of hepatocellular bile acid and organic anion uptake systems in *Xenopus laevis* oocytes. *Biochem J*, 316 (Pt 3):901–904, 1996.

[20] J L Boyer, O C Ng, M Ananthanarayanan, A F Hofmann, C D Schteingart, B Hagenbuch, B Stieger, and P J Meier. Expression and characterization of a functional rat liver Na+ bile acid cotransport system in COS-7 cells. *Am J Physiol*, 266:G382–387, 1994.

[21] G A Kullak-Ublick, J Glasa, C Böker, M Oswald, U Grützner, B Hagenbuch, B Stieger, P J Meier, U Beuers, W Kramer, G Wess, and G Paumgartner. Chlorambucil-taurocholate is transported by bile acid carriers expressed in human hepatocellular carcinomas. *Gastroenterology*, 113:1295–1305, 1997.

[22] Joachim Geyer, José R Godoy, and Ernst Petzinger. Identification of a sodium-dependent organic anion transporter from rat adrenal gland. *Biochem Bioph Res Co*, 316:300–306, 2004.

[23] S A Weinman, M W Carruth, and P A Dawson. Bile acid uptake via the human apical sodium-bile acid cotransporter is electrogenic. *JBC*, 273:34691–34695, 1998.

[24] Eric Y Zhang, Mitch A Phelps, Antara Banerjee, Chandra M Khantwal, Cheng Chang, Freek Helsper, and Peter W Swaan. Topology scanning and putative three-dimensional structure of the extracellular binding domains of the apical sodium-dependent bile acid transporter (SLC10A2). *Biochemistry*, 43:11380–11392, 2004.

[25] Daniel Zahner, Uta Eckhardt, and Ernst Petzinger. Transport of taurocholate by mutants of negatively charged amino acids, cysteines, and threonines of the rat liver sodium-dependent taurocholate cotransporting polypeptide Ntcp. *Eur J Biochem*, 270:1117–1127, 2003.

[26] Naissan Hussainzada, Tatiana Claro Da Silva, and Peter W Swaan. The cytosolic half of helix III forms the substrate exit route during permeation events of the sodium/bile acid cotransporter ASBT. *Biochemistry*, 48:8528–8539, 2009.

[27] Harini Krishnamurthy, Chayne L Piscitelli, and Eric Gouaux. Unlocking the molecular secrets of sodium-coupled transporters. *Nature*, 459:347–355, 2009.

[28] Dinesh Yernool, Olga Boudker, Yan Jin, and Eric Gouaux. Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature*, 431:811–818, 2004.

[29] Atsuko Yamashita, Satinder K Singh, Toshimitsu Kawate, Yan Jin, and Eric Gouaux. Crystal structure of a bacterial homologue of Na+/Cl-dependent neurotransmitter transporters. *Nature*, 437:215–223, 2005.

[30] Salem Faham, Akira Watanabe, Gabriel Mercado Besserer, Duilio Cascio, Alexandre Specht, Bruce A Hirayama, Ernest M Wright, and Jeff Abramson. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. *Science*, 321:810–814, 2008.

[31] Simone Weyand, Tatsuro Shimamura, Shunsuke Yajima, Shunichi Suzuki, Osman Mirza, Kuakarun Krusong, Elisabeth P Carpenter, Nicholas G Rutherford, Jonathan M Hadden, John O'Reilly, Pikyee Ma, Massoud Saidijam, Simon G Patching, Ryan J Hope, Halina T Norbertczak, Peter C J Roach, So Iwata, Peter J F Henderson, and Alexander D Cameron. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science*, 322:709–713, 2008.

[32] Carola Hunte, Emanuela Screpanti, Miro Venturi, Abraham Rimon, Etana Padan, and Hartmut Michel. Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. *Nature*, 435:1197–1202, 2005.

[33] B Stieger, B Hagenbuch, L Landmann, M Höchli, A Schroeder, and P J Meier. In situ localization of the hepatocytic Na+/Taurocholate cotransporting polypeptide in rat liver. *Gastroenterology*, 107(6):1781–1787, 1994.

[34] A B Hammann-Hänni. *Functional determinants of hepatocellular organic anion transporters: evolutionary comparison and domain analysis of two paralogs*. PhD thesis, University of Zurich, 2007.

[35] Toshimitsu Kawate and Eric Gouaux. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure*, 14:673–681, 2006.

[36] Jeroen R. Mesters and Rolf Hilgenfeld. Protein glycosylation, sweet to crystal growth? *Cryst Growth Des*, 7:2251–2253, 2007.

Chapter 3

Biophysical characterization of the extracellular loop of Oatp1a4

3.1 Introduction

3.1.1 Membrane transporters involved in drug disposition

The bioavailability of a drug is determined by several factors: Absorption, distribution, metabolism and excretion affect the ability of a drug to reach its target tissue. After oral administration of a substance, the efficiency of gastrointestinal uptake limits the absorption of the drug. In addition, the blood-brain barrier, the blood-placenta barrier or others protect tissues from exogeneous substances. Finally, compounds are actively removed from the body via hepatic clearance or elimination in the kidney.

Membrane transporters in the described tissues are responsible for the uptake of a substance or its extrusion. For the uptake of substances leading either to absorption of the compound or its elimination from the blood stream, peptide transporters (PEPTs), OATs and Oatps/OATPs are responsible. ABC transporters are the main extrusion system. Well-studied examples are P-glycoprotein or multidrug resistance proteins (MRPs).

Due to genetic variation, individuals react differently to the administration of the same drug. Pharmacogenetic studies addressing this issue are complicated by the diversity of transporters involved in uptake and elimination of drugs as well as their broad and overlapping substrate specificity. A detailed mechanistic understanding of the transport, the distribution and the regulation of the proteins executing the transport is therefore crucial for successful drug therapy.

3.1.2 The SLC21 superfamily of organic anion transporting polypeptides

Organic anion transporting polypeptides (human: OATPs, other species: Oatps) are multispecific transporters [1] that affect the pharmacokinetics of drugs [2]. To date, eleven human OATPs have been identified, and homologues are found in all mammals, frogs, *D. melanogaster* and *C. elegans* [3,4]. In yeast or bacteria, no homologs are known. Based on multiple sequence alignment, the proteins were assigned to the SLC21/SLCO superfamily and grouped in 6 families further divided in subfamilies [3].

OATPs/Oatps are expressed in the liver, the gastrointestinal tract, the kidney, the blood-brain barrier and the retina. Some transporters are found in a specific tissue. For instance, OATP1B1 and OATP1B3 are expressed only in the liver. Other OATPs/Oatps are ubiquitously expressed. In polarized cells, they reside in the basolateral membrane.

The substrate range of OATPs/Oatps is very broad. In general, large amphipathic molecules are transported. The specificity and transport kinetics of many OATPs/Oatps have been studied extensively (reviewed by [3,5,6]). Members of the Oatp1 family have a very broad substrate specificity: In addition to bile salt conjugates, it includes hormones and their conjugates, eicosanoids, peptides, various drugs and toxins. Table 3.1 summarizes the compounds transported by Oatp1a1, Oatp1a4 and Oatp2b1 from rat liver. Oatp1a4 was investigated in this study.

OATPs/Oatps act as anion exchangers. Translocation of the substrates is independent of the sodium gradient and electroneutral, and it is not clear whether transport is facilitative or secondary active. Little is known about the mechanistic details of transport. Glutathione and bicarbonate have been discussed as counterions for Oatp1a1 [7,8], while *cis*-inhibition or *trans*-stimulation of the transport by these ions was not observed for OATP1B1 and OATP1B3 [9]. In the same study, the electroneutral nature of transport was shown experimentally, thereby demonstrating the necessity of a counterion. It has been found that a low extracellular pH stimulates the transport activity and a histidine residue in transmembrane helix 3 has been identified that serves as a pH sensor [10].

The physiological role of OATPs/Oatps is the clearance of endogeneous and exogeneous compounds in liver and kidney, and the uptake of substances such as hormone conjugates into their target tissue. In addition to their role in drug disposition, OATPs/Oatps play a secondary role in the enterohepatic circulation of bile salts (section 2.1.2) as they contribute to the uptake of these molecules into hepatocytes.

Table 3.1: Substrate specificity of Oatps (modified from [1])

Protein Length MW	Tissue distribution	Substrates
Oatp1a1 670 AA 74.2 kDa	liver, kidney, brain, lung, retina, skeletal muscle, proximal colon	*Bile salts:* cholate, glycocholate, taurocholate, TCDCA, TUDCA, sulfotaurolithocholate *Hormones and hormone conjugates:* aldosterone, cortisol, DHEAS, $E_2 17\beta G$, estrone-3-sulfate, T_3, rT_3, T_4 *Peptides:* BQ-123, CRC220, deltorphin II, DPDPE, glutathione *Drugs:* dexamethasone, enalapril, fexofenadine, gadoxetate, ouabain, pravastatin, temocaprilat *Other organic anions:* monoglucuronosyl bilirubin, BSP, BSP-DSP-SG, E3040 glucuronide *Organic cations:* APD-ajmalinium, N-methylquinidine, rocuronium *Toxins:* ochratoxin A
Oatp1a4 661 AA 73.3 kDa	liver, kidney, brain, retina	*Bile salts:* cholate, glycocholate, taurocholate, TCDCA, TUDCA *Hormones and hormone conjugates:* DHEAS, $E_2 17\beta G$, estrone-3-sulfate, T_3, T_4 *Peptides:* BQ-123, DPDPE, Leu/enkephalin *Drugs:* biotin, digoxin, fexofenadine, ouabain, pravastatin *Organic anions:* APD-ajmalinium, rocuronium

Oatp2b1 682 AA 74.2 kDa	liver, lung, heart, brain, retina kidney	*Bile salts:* taurocholate *Eicosanoids:* LTC_4, prostaglandine E_1, prostaglandine E_2, prostaglandine D_2, thromboxane B_2 *Drugs:* iloprost

3.1.3 Structural features of OATPs/Oatps

OATPs/Oatps are glycoproteins of about 700 amino acids that share a twelve-transmembrane-domain topology with cytoplasmic termini, as predicted by hydropathy analysis. Figure 3.1 shows a topology model of Oatp1a4. While the structures of various ABC transporters including P-glycoprotein were determined by X-ray diffraction [11], experimental structural data is not available for any member of the SLC21 superfamily. A homology model was suggested by Meier-Abt et al. [4] based on the crystal structure of lactose permease. The model proposes a three-dimensional arrangement of the transmembrane helices with the connecting loops omitted. Lactose permease is a member of the major facilitator superfamily (MFS), one of the largest families of solute carriers [12]. Although the sequence similarity between OATPs/Oatps and MFS proteins is rather low, a recent study found indications that a large number of SLC families including SLC21 share the MFS fold [13].

The most characteristic structural feature found in all members of the SLC21 family is a large extracellular loop connecting transmembrane helix 9 and 10. This loop is likely to form a functionally important domain.

3.1.4 The extracellular loop of OATPs/Oatps

The extracellular loop between TM 9 and 10 of OATPs/Oatps comprises about 100 amino acids, including ten highly conserved cysteines. Hänggi *et al.* [14] used OATP2B1 to investigate the role of the domain. $\mathrm{OATP2B1}_{\Delta 489-557}$, a construct lacking the extracellular loop, was misprocessed and sequestered intracellularly. Individual point mutations of the cysteines to alanine resulted in reduced surface expression and decreased transport activity. The mutants could be shown to have a free thiol group, whereas the wild type did not, which led them to argue that all cysteines are likely to be involved in disulfide bridges. A distant homology to Kazal-type protease inhibitors has been suggested [1,4]. These small domains containing a short α-helix and a two-stranded β-sheet form several disulfide bonds. This finding further supports the hypothesis

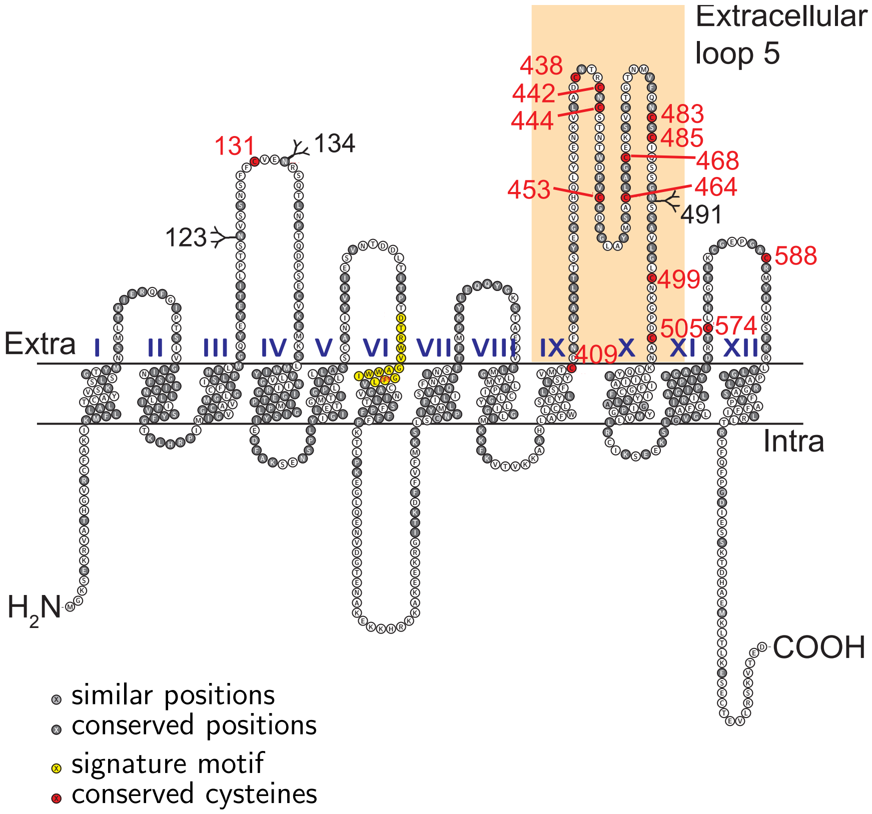


Figure 3.1: Topology of Oatp1a4, modified from [1]. The superfamily signature motif D-X-RW-(I,V)-GAWW-x-G-(F,L) is depicted in yellow. Gray shading shows conserved or similar positions. Conserved cysteine residues are colored red and their positions are indicated. Potential glycosylation sites labeled and drawn as trees. Extracellular loop 5 is highlighted.

that OATPs/Oatps contain a disulfide-rich domain. On the other hand, the cysteine pattern might also be involved in metal binding. Small metal-binding motifs such as zinc fingers often mediate protein-protein interactions.

3.1.5 Aim of the project

With the long-term goal to provide structural insight into OATPs/Oatps, we aimed to identify a fragment that would correspond to the extracellular domain of Oatp1a4. A soluble fragment of the protein would facilitate the elucidation of its structural and functional properties. We therefore designed and biophysically characterized several constructs of this domain.

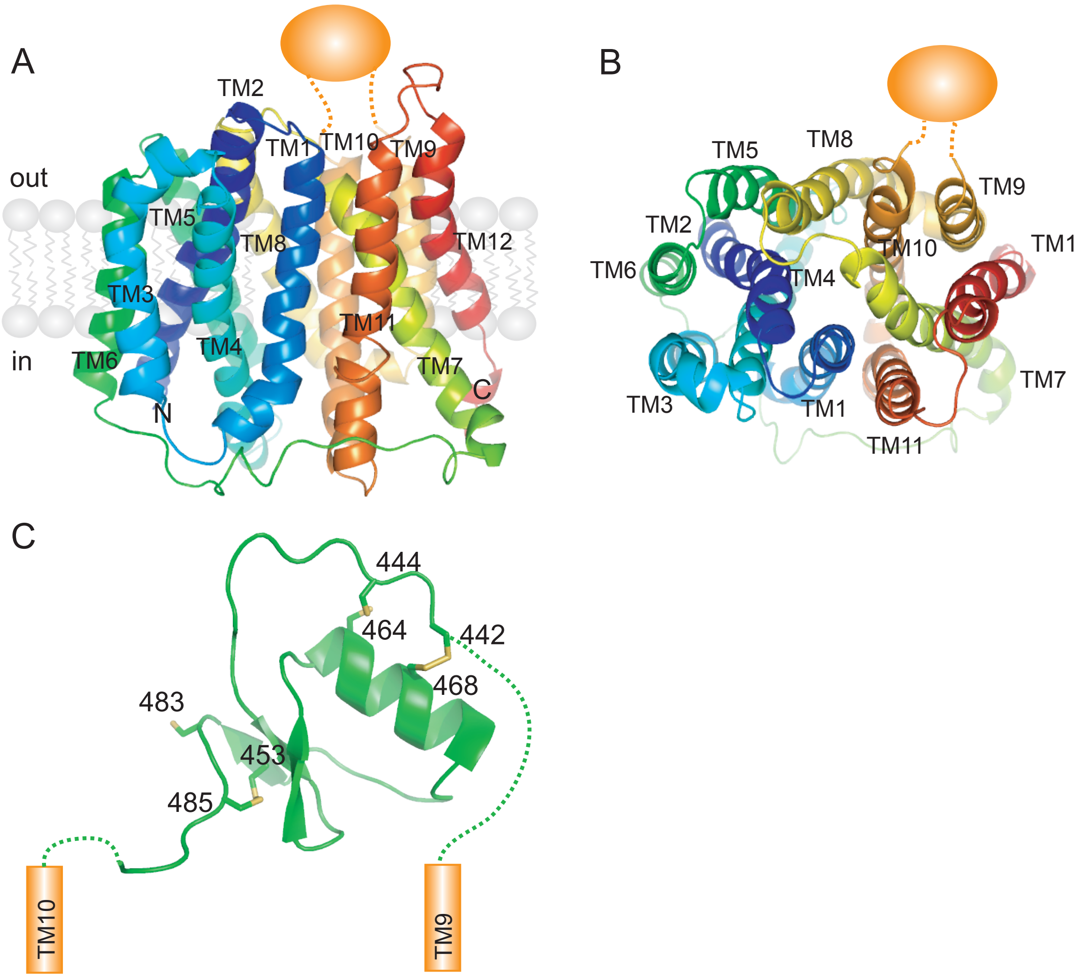


Figure 3.2: Computational models for Oatp1a4: (A) Model of the transmembrane region based on *Escherichia coli*EmrD (PDB ID 2GFP [15]). The extracellular domain is not part of the model. Its position is indicated with an orange sphere. (B) same model as seen from the extracellular side (rotated by 90 degrees towards the reader) (C) model of the extracellular domain of Oatp1a4 based on the structure of infestin 4 (PDB ID 2ERW [16]).

3.2 Results

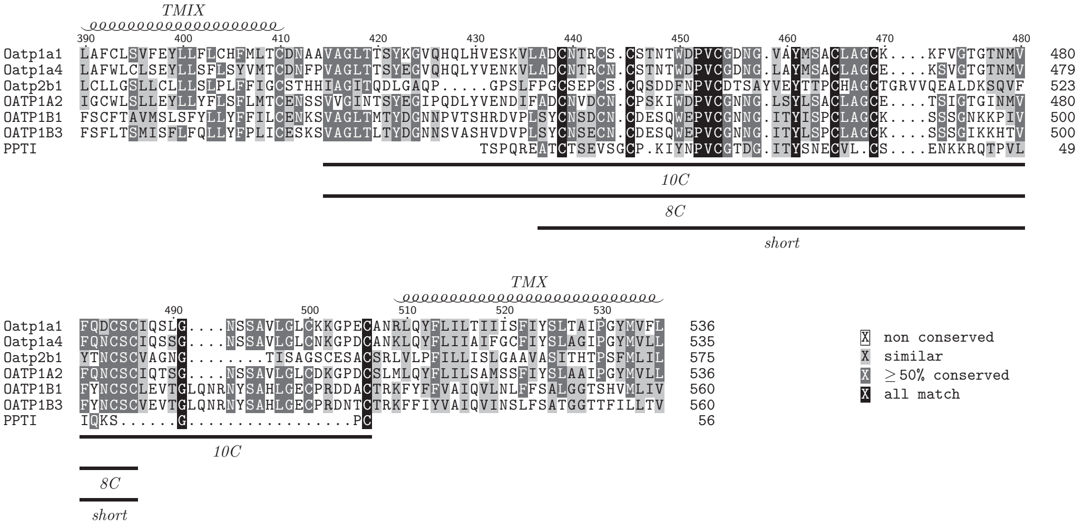
3.2.1 Model of Oatp1a4

We generated a model of Oatp1a4 to guide the design of expression constructs. The algorithm we used performs a search for the closest homolog with known structure based on the submitted sequence. For the full-length protein, a model showing the potential arrangement of the twelve transmembrane helices (Figure 3.2 A, B) is produced with the structure of the multidrug efflux protein EmrD from *Escherichia coli* [15] as a template. Similar models are obtained based on the structures of glycerol-3-phosphate transporter (GlpT) or lactose permease (LacY). All template proteins contain twelve transmembrane helices and belong to the major facilitator superfamily (MFS). The suggested structure corresponds to the previously published model of OATP1B3 [4] except that the position of TM1 and TM2 is swapped. In addition, our model includes most loops, which were omitted previously. The long intracellular connection between TM6 and TM7 lines the cytoplasmic part of the transporter.

However, the model lacks the residues representing the extracellular domain, located between TM9 and TM10. Hence, we submitted the corresponding sequence to the structure prediction server and obtained a model based on the crystal structure of the factor XIIa inhibitor infestin 4 [16], a serine protease inhibitor with a Kazal-type fold. The 48-residue model contains eight cysteines and suggests three disulfide bonds between residues 442 and 468, 444 and 464 as well as 453 and 485 (Figure 3.2 C).

3.2.2 Construct design

The design of constructs for the soluble expression of a fragment of Oatp1a4 was based on the hypothesis that extracellular loop 5 folds into a disulfide-rich domain homologous to Kazal type II serine protease inhibitors. Domain boundaries for our constructs were designed based on an alignment of six OATPs/Oatps taking into account the degree of conservation. We also aligned the sequences with porcine pancreatic trypsin inhibitor (Figure 3.3). Three constructs of different length were defined as follows: 10C corresponds to residues 414 to 505 of Oatp1a4, containing 92 amino acids and all ten conserved cysteines of the extracellular loop. 8C includes residues 414 to 485 (72 amino acids), whereas short consists of residues 436 to 485 of Oatp1a4 (50 amino acids). The latter two constructs have eight cysteines.



```
Oatp1a1  LAFCLSVFEYLLFLCHFMLTCDNAAVAGLTTSYKGVQHQLHVESKVLADCNTRCS.CSTNTWDPVCGDNG.VAYMSACLAGCK....KFVGTGTNMV  480
Oatp1a4  LAFWLCLSEYLLSFLSYVMTCDNFPVAGLTTSYEGVQHQLYVENKVLADCNTRCN.CSTNTWDPVCGDNG.LAYMSACLAGCE....KSVGTGTNMV  479
Oatp2b1  LCLLGSLLCLLLSLPLFFIGCSTHHIAGITQDLGAQP.....GPSLFPGCSEPCS.CQSDDFNPVCDTSAYVEYTTPCHAGCTGRVVQEALDKSQVF  523
OATP1A2  IGCWLSLLEYLLYFLSFLMTCENSSVVGINTSYEGIPQDLYVENDIFADCNVDCN.CPSKIWDPVCGNNG.LSYLSACLAGCE....TSIGTGINMV  480
OATP1B1  FSCFTAVMSLSFYLLYFFILCENKSVAGLTMTYDGNNPVTSHRDVPLSYCNSDCN.CDESQWEPVCGNNG.ITYISPCLAGCK....SSSGNKKPIV  500
OATP1B3  FSFLTSMISFLFQLLYFPLICESKSVAGLTLTYDGNNSVASHVDVPLSYCNSECN.CDESQWEPVCGNNG.ITYLSPCLAGCK....SSSGIKKHTV  500
PPTI                                              TSPQREATCTSEVSGCP.KIYNPVCGTDG.ITYSNECVL.CS....ENKKRQTPVL   49

Oatp1a1  FQDCSCIQSLG....NSSAVLGLCKKGPECANRLQYFLILTIIISFIYSLTAIPGYMVFL  536
Oatp1a4  FQNCSCIQSSG....NSSAVLGLCNKGPDCANKLQYFLIIAIFGCFIYSLAGIPGYMVLL  535
Oatp2b1  YTNCSCVAGNG........TISAGSCESACSRLVLPFILLISLGAAVASITHTPSFMLIL  575
OATP1A2  FQNCSCIQTSG....NSSAVLGLCDKGPDCSLMLQYFLILSAMSSFIYSLAAIPGYMVLL  536
OATP1B1  FYNCSCLEVTGLQNRNYSAHLGECPRDDACTRKFYFFVAIQVLNLFFSALGGTSHVMLIV  560
OATP1B3  FYNCSCVEVTGLQNRNYSAHLGECPRDNTCTRKFFIYVAIQVINSLFSATGGTTFILLTV  560
PPTI     IQKS......G.................PC                                56
```

Figure 3.3: ClustalW multiple alignment from TM9 to TM10 of six Oatps/OATPs with porcine pancreatic trypsin inhibitor. The domain boundaries of the three constructs 10C, 8C and short are indicated with bars below the alignment.

3.2.3 Soluble expression of the domain as a fusion protein

The three constructs were cloned into a variety of expression vectors summarized in Table 3.2. The three constructs 10C, 8C and short behaved very similarly with regards to expression in a given vector system. Since the extracellular domain of Oatp1a4 contains many cysteines, we first aimed at secreting the protein into the periplasm of *E. coli* in order to allow disulfide bond formation in an oxidizing environment. To this end, we tested a set of vectors containing a *pelB* leader peptide. A low level of expression was achieved only when the protein was fused to maltose binding protein (MBP). However, when using the pMAL vector system, the protein formed soluble aggregates and the fusion protein could not be cleaved proteolytically. In contrast, cytoplasmic production using a pET vector lead to high expression levels of the protein fused to NusA, thioredoxin, GB1 or MBP (Table 3.2). We used the construct for the cytoplasmic expression of the domain as an N-terminal fusion with MBP from a pET vector for further experiments. An additional construct was made in which the MBP-fusion protein is modified by a C-terminal *Strep*-tag.

Table 3.2: Vectors for the expression of the extracellular domain of Oatp1a4.

Vector	N-term. tag	Cleavage site	C-term. tag	expression
pET20b(+)	*pelB*		*His*	no expression
pMAL-pm1	*pelB*-MBP	Prescission		detectable
pMAL-pm1	*pelB*-MBP	Prescission	*His*	detectable
pGEX-6P1	GST	Prescission		no expression
pMAL-c2E	MBP	Ek		no expression
pETM80	periplasmic DsbC	TEV		no expression
pETM82	cytoplasmic DsbC	TEV		no expression
pETNus-1a	6xHis-NusA	TEV		expressed
pETTrx-1a	6xHis-thioredoxin	TEV		highly expressed
pETGB1-1a	6xHis-GB1	TEV		highly expressed
pETMBP-1a	6xHis-MBP	TEV		highly expressed
pETMBP-1a	6xHis-MBP	TEV	*Strep*	highly expressed

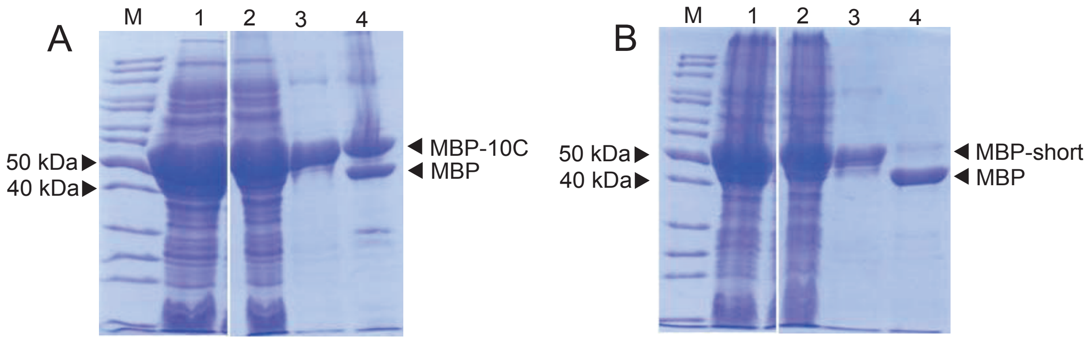


Figure 3.4: Purification of the extracellular domain fused to MBP. 12 % SDS PAGE of A) MBP-10C or B) MBP-short. Lane M) Size marker, 1) cell lysate, 2) flow-through IMAC, 3) elution IMAC, 4) TEV digest.

3.2.4 Properties of the purified extracellular loop

Purification of the fusion protein

Optimal expression of the domain with MBP fused to its N-terminus was obtained using LB medium after induction with 1 mM IPTG over night at 18 degrees. The cells were collected and lysed mechanically. The extracellular loop contains many cysteines that are kept reduced in the cytoplasm where the protein was produced. It proved to be crucial to include reducing agents in the lysis buffer to prevent protein precipitation upon exposure to oxygen. Both 1 mM DTT or a combination of 3 mM oxidized glutathion and 0.3 mM reduced glutathion were possible.

After lysis, the protein was purified via its *His*-tag. Alternatively, the MBP fusion protein could be purified by amylose affinity chromatography. While both methods yielded fusion protein of high purity, the Ni^{2+}-NTA agarose used for affinity purification via the *His*-tag was not compatible with DTT. Figure 3.4 shows a typical SDS-PAGE analysis of the affinity purification of MBP-10C and MBP-short.At this stage of purification, the yield from 1 l of expression culture was typically around 100 mg.

Aggregation behaviour of the fusion proteins

The aggregation behaviour of the fusion protein in various buffers was judged by gel filtration and is summarized in Table 3.3. Soluble aggregates were observed for MBP-10C and MBP-8C in 50 mM Tris/HCl pH 8, 150 mM NaCl without additives and in presence of various redox buffers such as different ratios of oxidized and reduced glutathion, L-cysteine and L-cystine, or DTT. The C-terminal *Strep*-tag also led to aggregation of the fusion protein. In general, MBP-10C and MBP-8C behaved similarly. In contrast, MBP-short existed in a monomer-dimer equilibrium when gel-filtrated in 50 mM Tris/HCl pH 8, 150 mM NaCl with or without added reducing agent. Under the same conditions, MBP isolated after cleavage of the fusion protein eluted in a single monomeric peak (Figure 3.5). Therefore, we conclude that the oligomerization behaviour was induced by short. We analyzed the fractions from the two peaks of MBP-short using non-reducing SDS PAGE. The dimer peak contains covalently linked and non-covalently linked MBP-short even in presence of 1 mM DTT. A smaller fraction of covalent higher-molecular-weight oligomers elutes in a shoulder. The presence of intermolecular links is an indication that the disulfide bonds are not formed correctly or completely upon oxidation of the protein.

Table 3.3: Aggregation behaviour of the fusion protein as judged by gel filtration in 50 mM Tris/HCl pH 8.0, 150 mM NaCl with the indicated additives. n.d. not determined

Condition	MBP-10C or MBP-8C	MBP-short
no additives	aggregation	monomer-dimer equilibrium
1 mM DTT	aggregation	monomer-dimer equilibrium
3 mM GSH, 0.3 mM GSSG	aggregation	monomer-dimer equilibrium
3 mM GSH, 1 mM GSSG	aggregation	n.d.
5 mM L-cysteine, 0.5 mM L-cystine	aggregation	n.d.
500 mM NaCl	aggregation	n.d.
0.2 % tween-20	aggregation	n.d.
Strep-tag	aggregation	aggregation

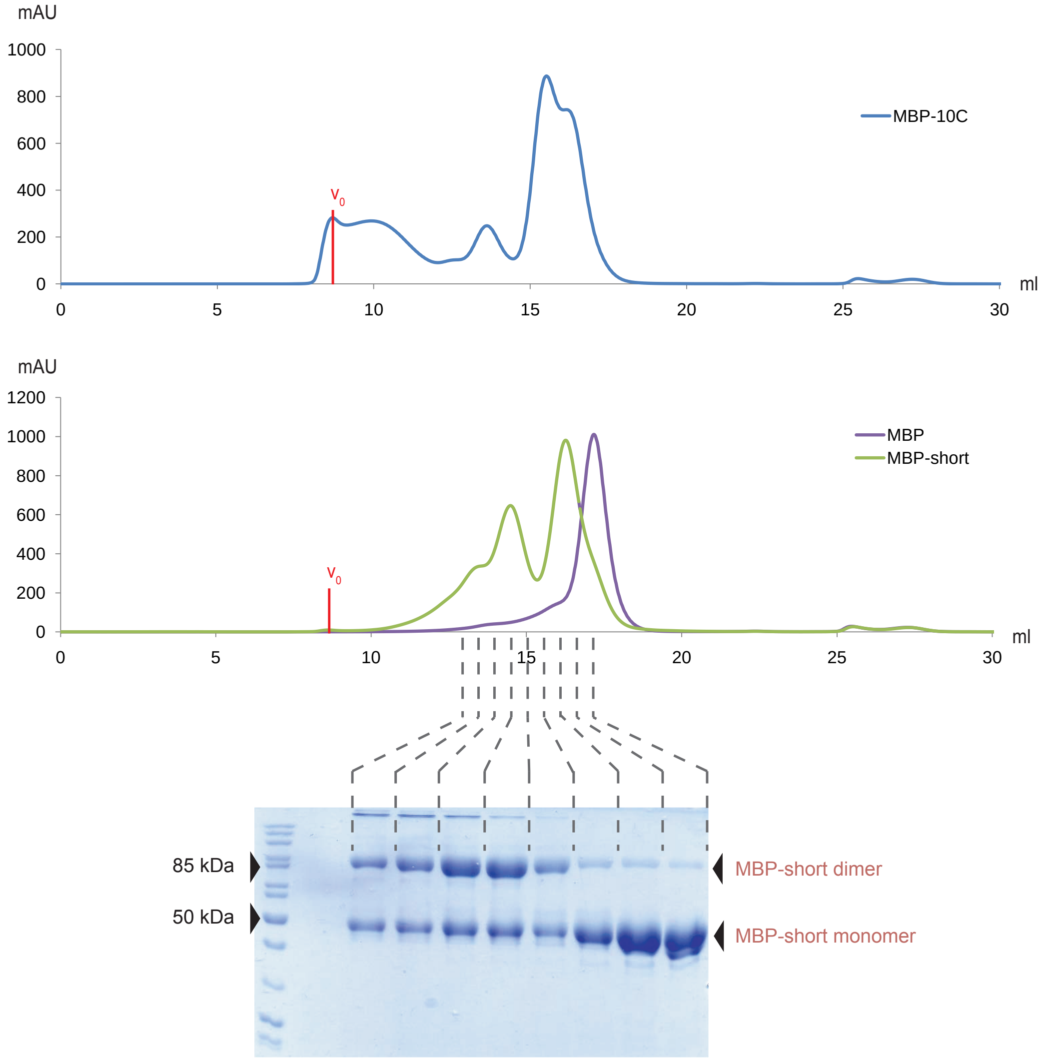


Figure 3.5: Gel filtration profile of MBP-10C, MBP-short and MBP in 50 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM DTT using a Superdex 200 column. The fusion protein exists in a monomer-dimer equilibrium, while MBP elutes in a monomeric peak. The indicated fractions of MBP-short were analyzed using SDS PAGE under non-reducing conditions.

Cleavage and isolation of the domain

All constructs used for expression contained a cleavage site for tobacco etch virus (TEV) protease at the N-terminus of the domain. After enzymatic cleavage, the fusion partner was removed *via* its N-terminal *His*-tag, while the domain was recovered with the flow-through. For constructs modified with a C-terminal *Strep*-tag, the domain could alternatively be captured by affinity chromatography using a streptavidin-coupled matrix. Since MBP-10C was prone to aggregation, proteolytic cleavage was not complete. About 50 % of the fusion protein was cleaved as judged by comparison of the intensity of the bands for MBP-10C and MBP when analyzed by SDS-PAGE (Figure 3.4 A). In contrast, MBP-short, which does not form high-molecular weight aggregates, is cleaved almost completely (Figure 3.4 B). Due to its small size, the domain could only be visualized by Coomassie staining when high amounts of protein were used for SDS-PAGE. Figure 3.4 C shows that the isolated domain was highly pure. Under non-reducing conditions, the domain migrates further during SDS-PAGE than under reducing conditions. This observation indicates that the domain is more compact in an oxidizing environment due to the presence of disulfide bonds. The typical yield from 1 l of expression culture was 1.5 mg.

Size exclusion chromatography was used as a polishing step and to judge the monodispersity of the protein. 10C elutes mainly in the void volume of a Superdex 75 column, while short is found in two main peaks at 12 and 14 ml elution volume. Suspecting that the aggregation behaviour is caused by incorrect formation of disulfide bonds, we used an approach to allow the protein to form the native interactions by dialysis in a redox buffer in presence of zinc. This would also enable the reconstitution of potential zinc-binding motifs. This disulfide shuffling protocol does not affect the properties of 10C. In short, however, the 12 ml peak significantly decreases in favour of the 14 ml peak.

The oligomeric state of the protein from the peak fraction at 14 ml was investigated in a sedimentation velocity analytical ultracentrifugation experiment. The analysis of the data showed a single species with a molecular mass of 7700 ± 1600 Da (Figure 3.7), which corresponds well to the theoretical mass of the short monomer of 6020.4 Da. The frictional ratio was fit to 1.29, which is a common value usually observed for globular proteins. The quality of the fit is high with a RMSD of 0.0076. We conclude that the 14 ml peak during size exclusion chromatography using a Superdex 75 column represents the monomer of purified short and used this fraction for all experiments described below.

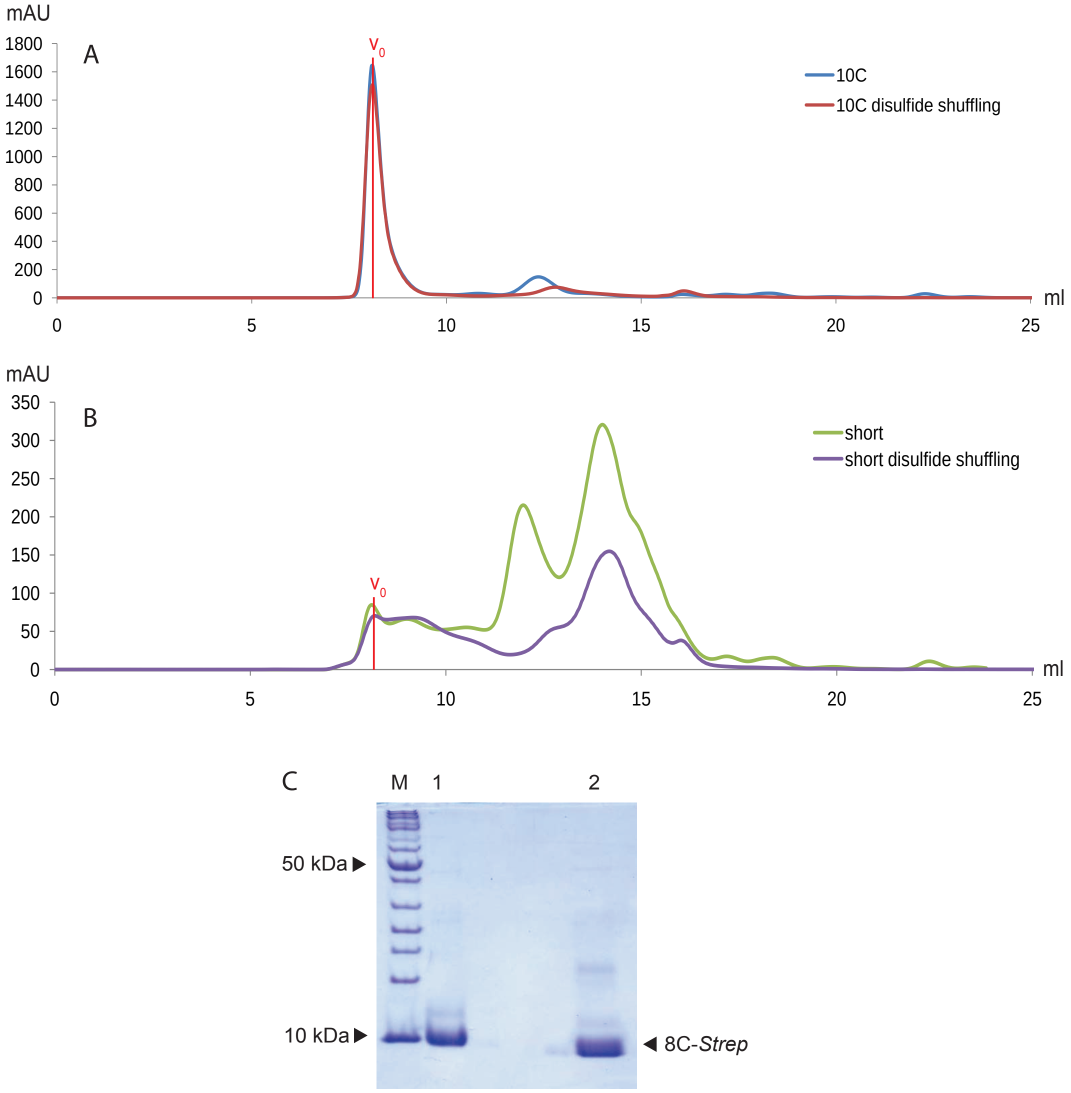


Figure 3.6: Gel filtration profile of (A) 10C or (B) short with or without disulfide shuffling in 50 mM Tris pH 8.0, 150 mM NaCl using a Superdex 75 column. (C) 15 % Schägger gel of purified 8C-*Strep*: M) Molecular weight marker 1) purified 8C-Strep under reducing conditions 2) purified 8C-Strep under non-reducing conditions. conditions.

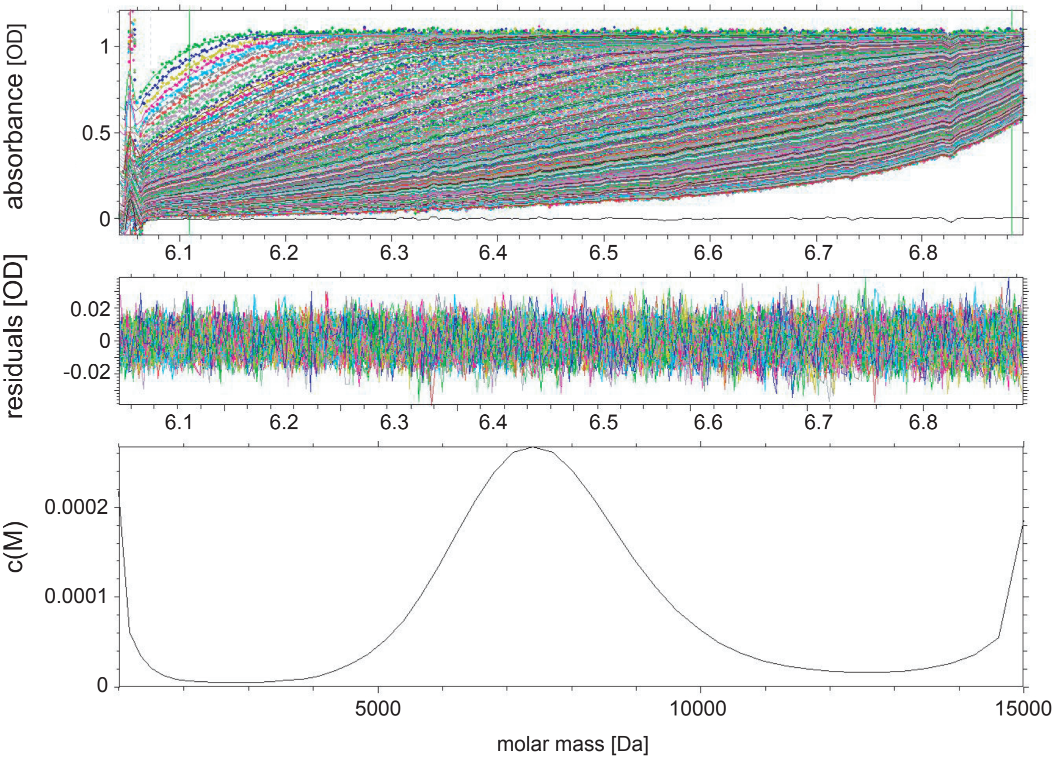


Figure 3.7: Sedfit analysis of sedimentation velocity analytical ultracentrifugation data for short. The absorption at 280 nm was monitored during centrifugation at 42000 rpm. The top panel shows the raw data. The residuals plotted in the middle panel indicate a good quality of the fit. The resulting mass is a single species with a molecular mass around 8000 Da.

Formation of disulfide bonds

In an effort to assess how many cysteine residues in the purified fragments are involved in disulfide bonds, free thiols were quantified using Ellman's reagent dithionitrobenzoic acid (DTNB). Each of the three fragments of Oatp1a4 had two free mercaptogroups. It is important to note that this value represents the average number of free thiols per molecule. The result was the same both under native and denaturing conditions, meaning that the free cysteines are accessible on the surface, thereby explaining the potential for intermolecular disulfide crosslinks.

We used the short construct in an attempt to identify these free cysteines. To this end, the protein was carbamidomethylated using iodoacetamide. The modification sites were then to be identified with ESI-MS. The unmodified domain has a mass of 6013.4 Da (Figure 3.8 A), while the predicted mass of short is 6020.4 Da. This finding agrees well with the observation that most cysteines are oxidized, since the formation of a disulfide bond results in a mass difference of 2 Da. When subjecting the domain to HPLC using a H_2O/acetonitrile gradient on a C8 column after treatment with iodoacetamide, the protein elutes in a broad peak at a retention time of 19-25 min (Figure 3.8 B). ESI-MS of the peak fraction shows a number of peaks with a mass difference of 56 Da (Figure 3.8 C), revealing that the sample contains unmodified short as well as protein modified with one, two, three, four or five carbamidomethyl groups. When the sample is reduced, the peaks corresponding to the domain with a different number of modifications can be recognized, but not seperated during HPLC. ESI-MS of the indicated fraction still shows a distribution of masses representing the domain with various numbers of carbamidomethyl groups (Figure 3.8 D). These results suggest that short does not have a defined structure, but forms disulfide bonds randomly when exposed to oxygen during lysis.

Random coil structure of the fragments

Since experiments described above suggested that the domain might be unstructured, we recorded CD spectra of both purified 10C and short in order to estimate the secondary structure content of the domain. In parallel, we wanted to investigate the effect of disulfide shuffling, as this treatment resulted in a different behaviour of the domain during gel filtration. The purified sample was devided in two parts. One part was left untreated, while the other was dialyzed in a redox buffer in presence of zinc to allow shuffling of erroneous disulfide bonds prior to the measurement. The absorption of polarized light was recorded from 250 to 190 nm for both constructs before or after redox shuffling (Figure 3.9). The secondary structure content was then estimated using SOM-CD (Table 3.4). While 10C contains 20 % alpha-helical and 40 % beta-sheet struc-

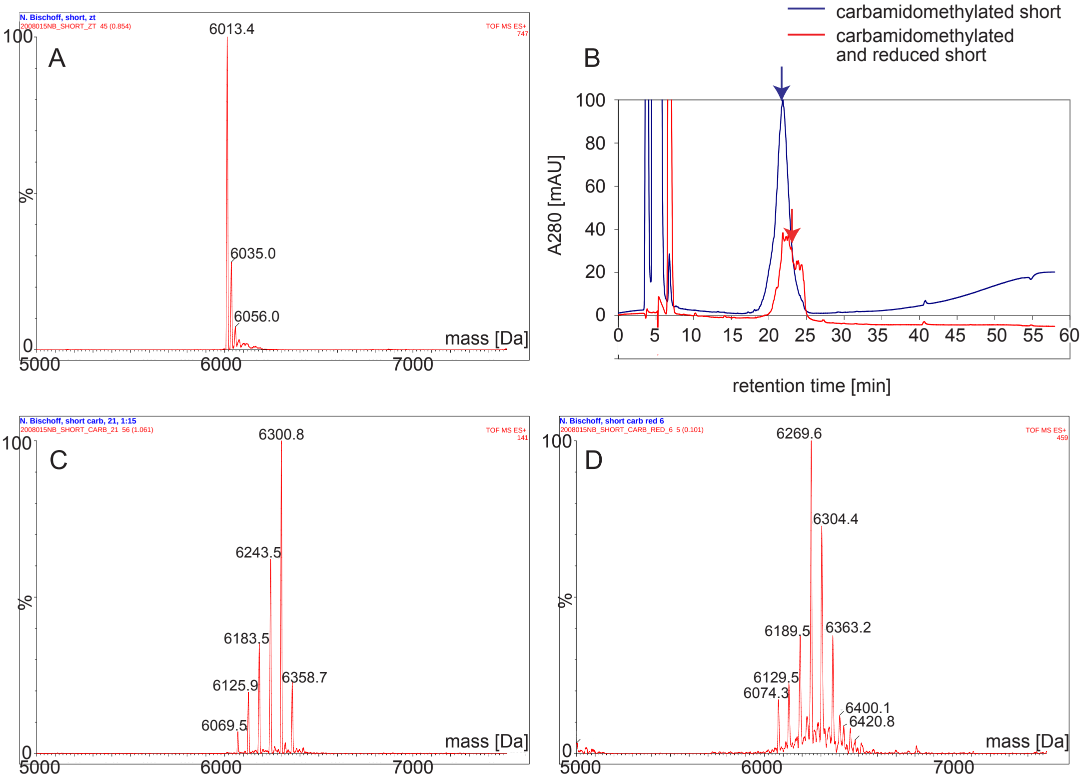


Figure 3.8: Analysis of carbamidomethylated short using HPLC and ESI-MS: (A) ESI-MS of purified short (B) HPLC of carbamidomethylated short before (blue) and after reduction (red). Fractions analyzed by ESI-MS are indicated by arrows in the respective color. (C) ESI-MS of carbamidomethylated short fractionated by HPLC (D) ESI-MS of carbamidomethylated and reduced short after fractionation by HPLC

ture, the CD spectrum of short indicates a random coil. Disulfide shuffling does not affect the secondary structure content of either domain. We conclude that the domain is not in its native state.

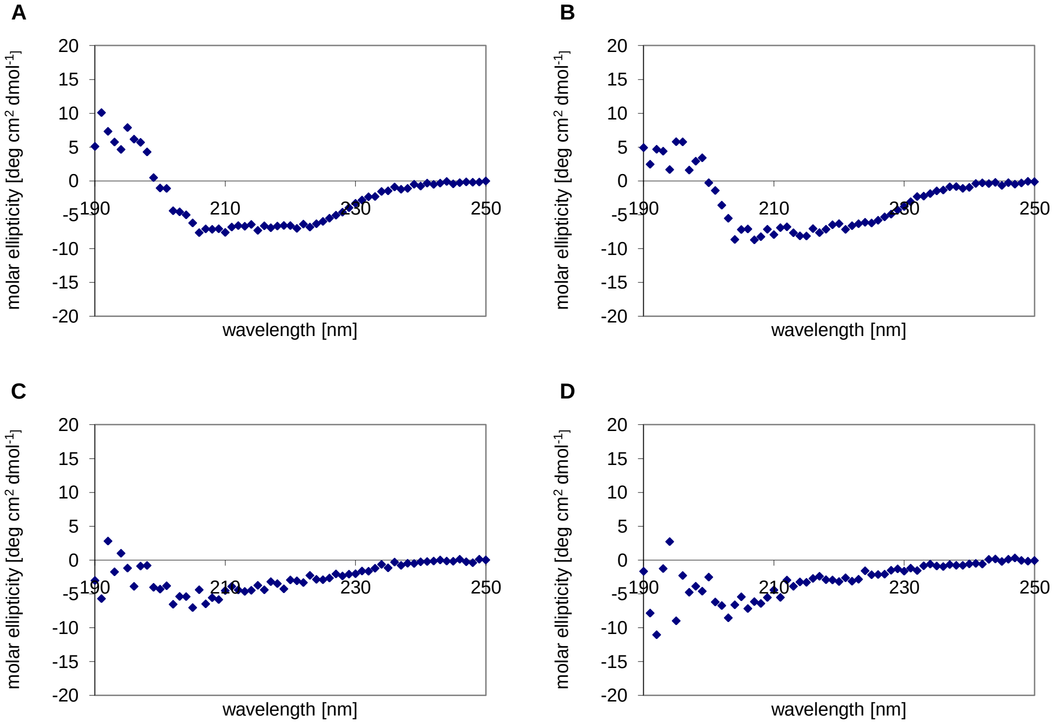


Figure 3.9: Circular dichroism spectra of A) 10C B) 10C after disulfide shuffling C) short D) short after disulfide shuffling. The spectra were recorded in 10 mM Tris/boric acid pH 8 at 4 degrees. The molar ellipticity is plotted versus the wavelength from 250 to 190 nm.

Table 3.4: Secondary structure content of domain estimated by SOM-CD

Construct	alpha (%)	beta (%)	turn (%)	random (%)
10C	20.2±9.9	41.5±4.4	9.8±1.2	28.5±4.8
10C after disulfide shuffling	20.2±9.9	41.5±4.4	9.8±1.2	28.5±4.8
short	1.3±4.5	2.6±9.1	1.1±4.0	95.0±17.6
short after disulfide shuffling	1.2±4.4	2.4±8.8	1.1±3.9	95.3±17.1

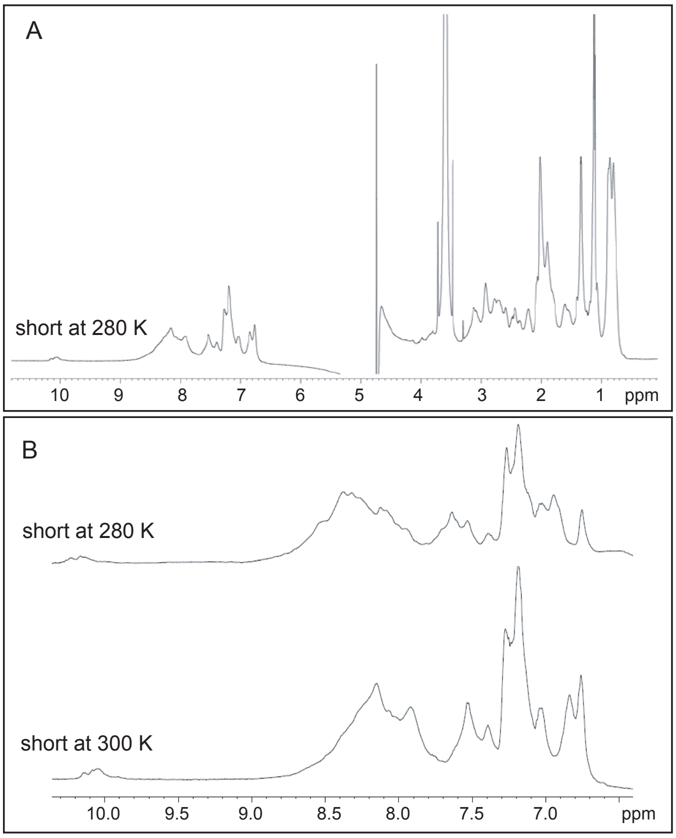


Figure 3.10: A) ^{1}H NMR spectrum of short at 280 K B) zoom on the amide region from 6.3 to 10.2 ppm at 280 K (top panel) or 300 K (bottom panel).

To investigate whether the domain is structured with an independent method, we recorded a 1D ^{1}H NMR spectrum of purified short at 280 K and 300 K (Figure 3.10). Random coil chemical shifts for main chain amide protons are typically observed at 8.1-8.8 ppm [17]. For a structured domain we would expect conformation-dependent chemical shift dispersion of the signals from backbone amide protons of folded regions above 8.5 ppm. However, the observed chemical shifts are typical for a random coil structure, showing only little dispersion in this region. An increase of temperature from 280 K to 300 K did not result in sharpening of the peaks, which is another indication for an unstructured protein sample.

3.3 Discussion

Due to the central role of transport proteins in drug absorption, distribution and extrusion, there is high interest in the study of the interaction of a potential therapeutic with the transporter. Organic anion transporting polypeptides (OATPs/Oatps) are important determinants of the bioavailability of drugs. Structural information on this protein class would therefore be of high value. One characteristic feature is the highly conserved extracellular loop 5 which connects transmembrane helices IX and X.

Aiming to produce a soluble domain using a bacterial expression system, we designed and analyzed constructs that correspond to the extracellular loop 5 of Oatp1a4 from *Rattus norvegicus*. We successfully expressed several fragments in high amounts in the cytoplasm of *E. coli* as a fusion protein with MBP that we purified to homogeneity. While longer constructs were polydisperse and aggregation-prone even in presence of the fusion partner, the shortest construct existed in an equilibrium between the monomer and a partly covalent dimer. Intramolecular crosslinking was observed even in presence of reducing agents. Due to its relatively large size, the fusion partner MBP likely prevented intermolecular disulfide bonds between more than two molecules.

Following cleavage of the fusion protein, we were able to isolate high amounts of the pure fragment. Longer constructs were highly aggregated, but seemed to contain secondary structure elements. In contrast, while the shortest construct was soluble, it had a random coil structure. Most cysteines were involved in disulfide bonds, but these covalent links did not follow a defined pattern. In place of a natively structured domain, the isolated protein consisted of a distribution of conformations with different non-native disulfide patterns as demonstrated by mass spectrometry following the modification of free cysteines.

The results suggest that the domain boundaries were not chosen correctly. The construct design had been based on the proposed homology of the extracellular loop of OATPs/Oatps to Kazal-type serine protease inhibitors [1,4]. Although this assumption is widely accepted in the field, the prediction of the fold of an extracellular domain of OATPs/Oatps might be biased by the occurence of a large number of cysteines leading to low complexity of the sequence. The neighboring extracellular loop 6 also contain several cysteine residues that may be in the vicinity of thiols in extracellular loop 5. Disulfide bonds between these residues may be present in the native structure of OATPs/Oatps. This hypothesis could be tested with a set of double mutants by detecting free cysteines with a sulfhydryl-reactive agent.

Cysteine-rich domains are common in cell-surface proteins. One example is the scavenger receptor cysteine-rich domain (SCRD), which is found at the extracellular C-terminus of scavenger receptors, but also in other cell surface proteins of immune cells [18]. The structures of several SRCR domains are known (e. g. [19] and show the presence of three disulfide bonds. Additional cysteines can be present that can be involved in chelating divalent cations, as observed for the SRCR domain of the macrophage receptor MARCO [20]. The domain mediates receptor-ligand interactions by domain swapping.

It is well known that cysteines do not exclusively form disulfide bonds, but may provide metal binding sites. Metal binding motifs such as zinc fingers are highly in soluble proteins, playing an important role in the interaction with DNA, but also in mediating protein-protein interaction [21]. Although the oligomeric state of OATPs/Oatps is not known, cross-linking experiments showed that the transporters are partly present in form of non-covalent dimers on the cell surface [14]. Metal-binding motifs are also known in membrane proteins: Cysteine-loop receptors such as the pentameric contain a low-affinity and a high-affinity regulatory zinc binding site located in a loop at the interface between the transmembrane part and the soluble domain of the protein.

A previous study showed that substitution of each of the cysteines of the extracellular loop 5 with alanine resulted in a free thiol group that could be detected with a sulfhydryl-reactive reagent [14]. This observation suggests that all cysteines are involved in disulfide bonds. However, a metal ion bound with high affinity would also shield cysteines from thiol-reactive agents.

As long as a soluble domain cannot be produced that adopts the native conformation of the extracellular loop 5 in OATPs/Oatps, we have to rely on information from the full-length protein. Assays to measure uptake of [^{3}H]-labeled substrate into CHO cells stably transfected with OATPs/Oatps are established [22]. We performed initial experiments to investigate the effect of zinc or chelating agents on the transport activity of rat liver Oatps. Here, conditions need to be found that do not affect the viability of the cells.

3.4 Experimental procedures

Materials used for this project are listed in the appendix.

3.4.1 Protein structure prediction

The homology models of Oatp1a4 and the extracellular loop were generated by submitting the sequence of the full-length protein as well as residues 410-507, which correspond to the extracellular loop 5, to the structure prediction program Phyre [23]. This algorithm uses profile-profile matching based on predicted secondary structure to build a three-dimensional model.

3.4.2 Cloning of the extracellular loop of Oatp1a4

The DNA sequence of the extracellular loop corresponding to residues 414 to 505, 414 to 485 or 436 to 485 of Oatp1a4 was amplified by PCR from the template pSPORT-Oatp1a4 [24] to yield the constructs 10C, 8C or short, respectively. The resulting products were cloned into pET20b(+), pMAL-pm1, pMAL-c2E pGEX-6P1, pETM80, pETM82, pETNus-1a, pETTrx/1a, pETGB1-1a and pETMBP-1a. Constructs in pET20b(+), pMAL-pm1 and pETM80 were designed for secretion to the periplasm, whereas the other constructs were for protein expression in the cytoplasm. The described expression vectors allowed the screening of various N-terminal fusions of the constructs with solubility enhancers such as MBP, GST, GB1, thioredoxin or NusA. All N-terminal tags were followed by a sequence encoding for a specific cleavage site for Prescission, enterokinase or tobacco etch virus (TEV) protease.

3.4.3 Expression and purification of the domain

Small-scale expression tests

Expression tests were performed in 5 ml cultures in LB medium. A single colony of *Escherichia coli* BL21(DE3) freshly transformed with the expression vector was used to inoculate 5 ml LB containing 100 μg/ml ampicillin or 50 μg/ml kanamycin. The starter culture was incubated over night at 37 degrees and 230 rpm. The next morning, 100 μl of the stationary culture were used to inoculate 5 ml fresh medium. Expression was induced at $\mathrm{OD}_{600} = 0.5$ by addition of IPTG to 1 mM for 3 hours at 37 degrees and 230 rpm or at 18 degrees overnight at 230 rpm. Samples were taken before and after induction and analyzed by SDS PAGE.

Expression and purification of the MBP fusion protein

MBP-10C, MBP-8C or MBP-short with or without C-terminal *Strep*-tag were expressed from pETMBP-1a. A single colony of *Escherichia coli* strain BL21(DE3) freshly transformed with the expression vector was used to inoculate 5 ml LB containing 50 μg/ml kanamycin. The starter culture was grown over night at 37 degrees and 230 rpm. The next day,500 ml of LB containing 50 μg/ml kanamycin in a 2 l baffled flask were inoculated with the starter culture and incubated at 37 degrees and 110 rpm. Expression was induced at $\mathrm{OD}_{600} = 0.8$ with 1 mM IPTG over night at 18 degrees. After harvesting the cells by centrifugation for 15 min at 4000 rpm and 4 degrees, the cell pellets were flash-frozen in liquid nitrogen and stored at -80 degrees until use.

For purification, we typically used cells from 1 l of expression culture. The pellets were thawed on ice in 20 ml lysis buffer (50 mM Tris/HCl pH 8.0, 150 mM NaCl, 3 mM GSH, 3 mM GSSG, 0.1 mg/ml DNAseI or 50 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM DTT, 0.1 mg/ml DNAseI). The cells were disrupted mechanically by three rounds of French press at 4 degrees. After removing cell debris by centrifugation for 20 min at 5000 rmp and 4 degrees, the supernatant was subjected to affinity chromatography using Ni-NTA agarose (Qiagen) or amylose resin (New England Biolabs).

To isolate the domain, the purified fusion protein was cleaved using 10 μg of tobacco etch virus (TEV) protease per mg of protein during dialysis in 50 mM Tris pH 8.0, 150 mM NaCl, 3 mM GSH, 0.3 mM GSSG (or 1 mM DTT) over night at 4 degrees. The domain was isolated via its C-terminal *Strep*-tag during affinity chromatography using StepTactin (IBA Biotech) or by passing the mixture three times over a Ni-NTA column equilibrated with 50 mM Tris pH 8.0, 150 mM NaCl, 10 mM imidazol. The domain is recovered from the flow-through, while the His-tagged fusion partner, uncleaved protein and the protease are bound to the column material.

3.4.4 Size exclusion chromatography

Size exclusion chromatography was used as a polishing step following purification, but also to judge the monodispersity of the protein preparation. The fusion protein was analyzed using a Superdex 200 10/300 GL column (GE Healthcare), while preparative and analytical gel filtrations were performed on a Superdex 75 10/300 GL column (GE Healthcare).

3.4.5 Ellman assay

The reaction of dithionitrobenzoic acid (DTNB) with accessible thiols goes along with the stoichiometric release of a nitrothiobenzoate anion. The yellow compound can easily be quantified by absorption spectroscopy. The so-called Ellman's reagent is commonly used to determine the number of free cysteines in a protein sample.

We used N-acetylcysteine to prepare a standard curve. Triplicates of 1 ml samples of 0, 5, 10, 15, 20, 30, 40 and 50 $/\mathrm{muM}$ N-acetylcysteine in the reaction buffer (50 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM EDTA) were mixed with 50 μl of 3 mM DTNB and incubated at 37 degrees for 30 min. The absorbance was then measured at 412 nm in a spectrometer using the buffer as a reference. The concentration of N-acetylcysteine was plotted versus the average absorption value and a linear curve fit was applied. To determine the number of free cysteines in the native sample, 200 μl of the protein solution with known concentration (5-10 μM) was diluted 1:5 in the reaction buffer. After addition of 50 μl 3 mM DTNB, the samples were incubated for 30 min at 37 degrees and the absorption was measured at 412 nm. For the denatured sample, the procedure was repeated diluting the protein in 8 M urea in reaction buffer. To quantify the total number of cysteines, 100 μl of the protein solution were diluted 1:5 in 125 mM DTT, 8 M urea in reaction buffer and reduced for 30 min at 30 degrees. DTT was then removed by buffer exchange using a NAP-5 column equilibrated with 6.4 M urea in reaction buffer. 1 ml of the eluate was mixed with 50 μl of 3 mM DTNB after determination of the protein concentration. After incubation for 30 min at 37 degrees, the absorbance at 412 nm was determined. The respective number of cysteines per molecule was calculated using the standard curve.

3.4.6 Modification of free cysteines

Iodoacetamide is a nucleophile that reacts with free cysteines, resulting in the modification of the side chain with a carbamidomethyl group. We treated purified short with iodoacetamide to identify cysteine residues not involved in disulfide bonds. To this end, 0.5 volume of 1 M iodoacetamide was added to purified and concentrated short (3-4 mg/ml). After allowing the reaction to proceed for 30 min at room temperature in the dark, the reaction was quenched by adding 1/100 volume of 10 % trifluoroacetic acid (TFA). The sample was analyzed by HPLC using a C8 column with a H_2O/acetonitrile gradient. Fractions were collected manually and analyzed by ESI-MS using the service of the Functional Genomics Center Zurich. For reduction of disulfide bonds, fractions were lyophilized, resuspended in 100 μl 50 mM Tris/HCl pH 8.0, 150 mM NaCl and mixed with 50 μl of 1 M DTT. After incubation at room temperature for 3 h, 100 μl were used for HPLC analysis and subsequent ESI-MS.

3.4.7 Analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation to determine the oligomeric state of the purified domain were performed in an Optima XL-1 centrifuge (Beckman Coulter) with a Ti50 rotor using epoxy counter pieces with sapphire windows. After gel filtration, the protein sample with $\mathrm{A}_2 80 \approx 1$ was filled in the sample chamber, while the reference cell was loaded with gel filtration buffer. The absorption at 280 nm was measured during centrifugation at 42000 rpm at 20 degrees. The data were analyzed using Sedfit [25].

3.4.8 Circular dichroism

We recorded circular dichroism (CD) spectra using a Jasco-715 spectrometer (Jasco, Japan). The protein sample was dialyzed three times in 50 volumes of 10 mM Tris/boric acid pH 8 at 4 degrees. For the measurement, the protein was diluted to 10 μM and filled into a cuvette with a path length of 1 mm. The CD signal in mdeg was scanned from 250 to 190 nm in 1 nm increments. After substraction of the reference sample, the signal was converted to molar ellipticity and the secondary structure content was estimated using SOM-CD [26].

3.4.9 NMR

Nuclear magnetic resonance (NMR) was carried out by Prof. Oliver Zerbe at the Department of Organic Chemistry of the University of Zurich in order to judge whether short forms a structured domain. One-dimensional ^{1}H NMR spectra of 500 μM purified short in 20 mM Tris/HCl pH 8.0, 50 mM NaCl were recorded at 280 K and 300 K.

Bibliography

[1] B Hagenbuch and P J Meier. The superfamily of organic anion transporting polypeptides. *BBA*, 1609(1):1–18, 2003.

[2] A Kalliokoski and M Niemi. Impact of OATP transporters on pharmacokinetics. *Brit J Pharmacol*, 158(3):693–705, 2009.

[3] Bruno Hagenbuch and Peter J Meier. Organic anion transporting polypeptides of the OATP/ SLC21 family: phylogenetic classification as OATP/ SLCO superfamily, new nomenclature and molecular/functional properties. *Pflug Arch Eur J Phy*, 447(5):653–65, 2004.

[4] Fabienne Meier-Abt, Younes Mokrab, and Kenji Mizuguchi. Organic anion transporting polypeptides of the OATP/SLCO superfamily: identification of new members in nonmammalian species, comparative modeling and a potential transport mode. *J Membrane Biol*, 208(3):213–227, 2005.

[5] B Hagenbuch and C Gui. Xenobiotic transporters of the human organic anion transporting polypeptides (OATP) family. *Xenobiotica*, 38:778–801, 2008.

[6] Jörg König, Annick Seithel, Ulrike Gradhand, and Martin F Fromm. Pharmacogenomics of human OATP transporters. *N-S Arch Pharmacol*, 372:432–443, 2006.

[7] L Li, T K Lee, P J Meier, and N Ballatori. Identification of glutathione as a driving force and leukotriene c4 as a substrate for oatp1, the hepatic sinusoidal organic solute transporter. *JBC*, 273(26):16184–16191, 1998.

[8] L M Satlin, V Amin, and A W Wolkoff. Organic anion transporting polypeptide mediates organic anion/HCO3- exchange. *JBC*, 272(42):26340–26345, 1997.

[9] Chitrawina Mahagita, Steven M Grassl, Pawinee Piyachaturawat, and Nazzareno Ballatori. Human organic anion transporter 1B1 and 1B3 function as bidirectional carriers and do not mediate GSH-bile acid cotransport. *Am J Physiol-Gastr L*, 293(1):G271–278, 2007.

[10] Simone Leuthold, Bruno Hagenbuch, Nilufar Mohebbi, Carsten A. Wagner, Peter J. Meier, and Bruno Stieger. Mechanisms of pH-gradient driven transport mediated by organic anion polypeptide transporters. *Am J Physiol Cell Physiol*, 296(3):C570–582, 2009.

[11] Stephen G Aller, Jodie Yu, Andrew Ward, Yue Weng, Srinivas Chittaboina, Rupeng Zhuo, Patina M Harrell, Yenphuong T Trinh, Qinghai Zhang, Ina L Urbatsch, and Geoffrey Chang. Structure of p-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, 323:1718–1722, 2009.

[12] S S Pao, I T Paulsen, and M H Saier. Major facilitator superfamily. *Microbiol Mol Biol R*, 62(1):1–34, 1998.

[13] Avner Schlessinger, Pär Matsson, James E Shima, Ursula Pieper, Sook Wah Yee, Libusha Kelly, Leonard Apeltsin, Robert M Stroud, Thomas E Ferrin, Kathleen M Giacomini, and Andrej Sali. Comparison of human solute carriers. *Protein Science*, 19(3):412–428, 2010.

[14] Emanuel Hänggi, Anne Freimoser Grundschober, Simone Leuthold, Peter J Meier, and Marie V St-Pierre. Functional analysis of the extracellular cysteine residues in the human organic anion transporting polypeptide, OATP2B1. *Mol Pharmacol*, 70(3):806–17, 2006.

[15] Yong Yin, Xiao He, Paul Szewczyk, That Nguyen, and Geoffrey Chang. Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science*, 312:741–744, 2006.

[16] J A R G Barbosa, I T N Campos, and A S Tanaka. Structural studies of infestin 4, a factor XIIa inhibitor. *Acta cryst*, A61:C204–C205, 2005.

[17] Kurt Wüthrich. *NMR of proteins and nucleic acids*. John Wiley & sons, Inc., 1986.

[18] Maria Rosa Sarrias, Jørn Grønlund, Olga Padilla, Jens Madsen, Uffe Holmskov, and Francisco Lozano. The scavenger receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. *Crit Rev Immunol*, 24(1):1–37, 2004.

[19] E Hohenester, T Sasaki, and R Timpl. Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. *Nat Struct Biol*, 6(3):228–232, 1999.

[20] Juha R M Ojala, Timo Pikkarainen, Ari Tuuttila, Tatyana Sandalova, and Karl Tryggvason. Crystal structure of the cysteine-rich domain of scavenger receptor MARCO reveals the presence of a basic and an acidic cluster that both contribute to ligand recognition. *JBC*, 282(22):16654–16666, 2007.

[21] Roland Gamsjaeger, Chu Kong Liew, Fionna E Loughlin, Merlin Crossley, and Joel P Mackay. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.*, 32(2):63–70, 2007.

[22] U Eckhardt, A Schroeder, B Stieger, M Höchli, L Landmann, R Tynes, P J Meier, and B Hagenbuch. Polyspecific substrate uptake by the hepatic organic anion transporter oatp1 in stably transfected CHO cells. *Am J Physiol*, 276:G1037–1042, 1999.

[23] Lawrence A Kelley and Michael J E Sternberg. Protein structure prediction on the web: a case study using the phyre server. *Nat Protoc*, 4(3):363–371, 2009.

[24] B Noé, B Hagenbuch, B Stieger, and P J Meier. Isolation of a multispecific organic anion and cardiac glycoside transporter from rat brain. *PNAS*, 94(19):10346–10350, 1997.

[25] Patrick H Brown and Peter Schuck. A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Comput. Phys. Commun.*, 178:105–120, 2008.

[26] P Unneberg, J J Merelo, P Chacón, and F Morán. SOMCD: method for evaluating protein secondary structure from UV circular dichroism spectra. *Proteins*, 42(4):460–470, 2001.

Chapter 4

Designed ankyrin repeat protein binders for the crystallization of AcrB: Plasticity of the dominant interface

Nicole Monroe[1], Gaby Sennhauser[1,2], Markus Seeger[1], Christophe Briand[1], Markus G. Grütter[1*]

[1]Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland
[2]Present address: Molecular Partners AG, Wagistrasse 14, 8952 Zurich-Schlieren, Switzerland

*Corresponding author

Journal of Structural Biology, 174(2):269-81, 2011.

Author contributions: Nicole Monroe received the DARPin libraries after ribosome selection to AcrB from Gaby Sennhauser. She determined two new crystal structures of AcrB-DARPin complexes with the help of Christophe Briand and carried out all experiments to biophysically characterize the interaction. Nicole Monroe wrote the manuscript and prepared all figures with helpful corrections from Markus Seeger and Markus Grütter.

4.1 Abstract

The formation of well-diffracting crystals is a major bottleneck in structural analysis of membrane proteins by X-ray crystallography. One approach to improve crystal quality is the use of DARPins as crystallization chaperones. Here, we present a detailed analysis of the interaction between DARPins and the integral membrane protein AcrB. We find that binders selected *in vitro* by ribosome display share a common epitope. The comparative analysis of three crystal structures of AcrB-DARPin complexes allowed us to study the plasticity of the interaction with this dominant binding site. Seemingly redundant AcrB-DARPin crystals show substantially different diffraction quality as a result of subtle differences in the binding geometry. This work exemplifies the importance to screen a number of crystallization chaperones to obtain optimal diffraction data. Crystallographic analysis is complemented by biophysical characterization of nine AcrB binders. We observe that small variations in the interface can lead to differing behavior of the DARPins with regards to affinity, stoichiometry of the complexes and specificity for their target.

4.2 Introduction

While membrane proteins are important drug targets, high-resolution structural information on this protein class is still difficult to obtain. In the solubilized state, the detergent micelle shields the hydrophobic surface of the protein that would naturally be embedded in the membrane. This area is therefore not amenable to forming crystal contacts. The hydrophilic surface of a membrane protein, often consisting of loops connecting the membrane-spanning secondary structure elements, is comparatively small. One strategy to mediate crystal contacts is the use of soluble binding proteins to provide a larger surface for intermolecular interaction.

An additional complication is that transporters and receptors require intrinsic flexibility to carry out their function. Conformational heterogeneity, however, is not favorable for crystallization as it increases the entropic cost of lattice formation. Stabilization of a dynamic part of the crystallization target or shifting a thermodynamic equilibrium to one conformation can be achieved by a ligand or inhibitor. A prominent example is the use of vanadate to trap ABC transporters in a nucleotide-bound conformation [1]. Natural or designed proteinaceous binding partners are an attractive alternative.

Traditionally, antibody fragments were used to improve the quality of membrane protein crystals [2]. One example is the potassium channel KcsA crystallized in complex with Fab fragments [3], yielding a structure at 2 Å resolution as compared to 2.8 Å without crystallization chaperone. The antibody fragment mediates all crystal contacts and could be used for phasing by molecular replacement. In addition to antibody-based binding proteins, alternative scaffolds have become more widely used [4].

Designed ankyrin repeat proteins (DARPins) are modular proteins for the evolutionary design of protein binders [5]. Structurally, a repeat is composed of an alpha-turn-alpha motif followed by a loop that connects to the next repeat [5,6]. The signature motif TPLH contains a proline preceding the first helix which induces a kink at the transition from the loop to the helix, resulting in an overall L-shape of the module. The lateral surfaces of the repeats are hydrophobic, allowing the assembly of the modules in a stacking mode. Internal repeats are based on a consensus sequence, and are flanked by an N-terminal and a C-terminal capping module with a charged surface, thus preventing aggregation. As for the nomenclature, NxC is a DARPin with x internal repeats flanked by the terminal caps.

Libraries have been constructed based on the 33 amino acid consensus design [5]. Six residues in the potential protein-protein interaction interface are allowed to be any amino acid except glycine, proline or cysteine. An additional degenerated codon allows for the incorpora-

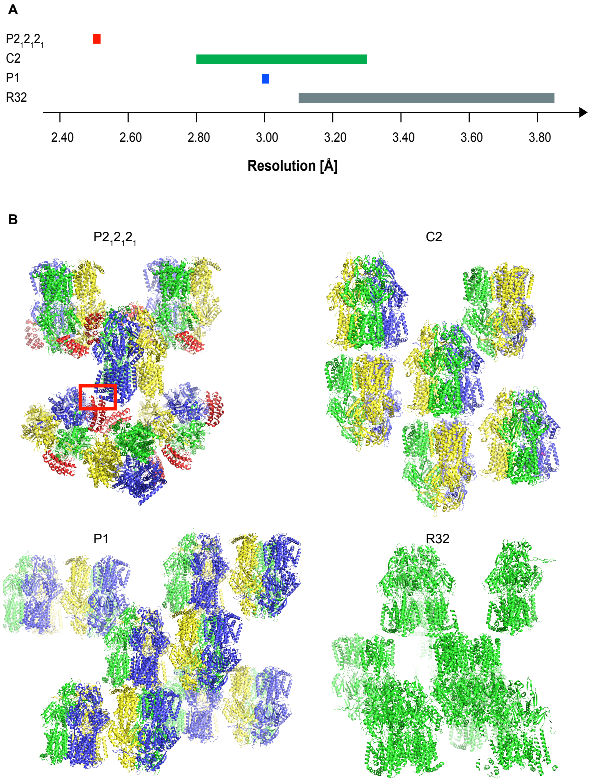


Figure 4.1: (A) Resolution of AcrB structures in space groups $P2_12_12_1$, C2, P1 and R32 published in the PDB (B) Packing of AcrB crystals: In space groups $P2_12_12_1$ and C2, there is one AcrB trimer or AcrB-DARPin complex in the asymmetric unit, while in case of P1, there are two trimers per asymmetric unit. The subunits of AcrB are shown in green, yellow and blue. DARPins are colored red. In the crystal form with space group R32, the protomers of the AcrB trimer (green) are related by crystallographic symmetry. The area marked with a red box is shown in detail in Figure 4.6.

tion of asparagines, tyrosine or histidine at position 26, the last residue of the second helix. In an N3C format containing three internal repeats flanked by capping modules, this randomization results in a theoretical diversity of $3.8 \cdot 10^{23}$. In contrast to antibodies and antibody fragments, DARPins are cysteine-free proteins characterized by extraordinary thermodynamic and chemical stability as well as high expression yields [5].

Ribosome display [7] and phage display [8] have been successfully used to select binders to a variety of protein targets including membrane proteins [9–11], and structures of many of the resulting complexes proof the suitability of DARPins as crystallization chaperones [10,12–18].

The first membrane protein target that yielded a crystal structure in complex with a DARPin crystallization chaperone is AcrB from *Escherichia coli*. AcrB belongs to the superfamily of RND (resistance/nodulation/cell division) transporters [19] and is the inner-membrane component of the AcrAB-TolC multidrug efflux system. The proton-driven transporter has been extensively characterized on a genetic, structural and functional level [20]. To date, 26 crystal structures of the protein have been deposited to the PDB (Figure 4.1 A). The first structure was solved in space group R32 and showed a symmetric conformation of the homotrimer with the three-fold symmetry axis relating the protomers [21]. The best resolution that could be obtained in this crystal form was 2.7 Å [22]. Later, AcrB was crystallized in an asymmetric form in space groups C2, P1 [23,24] and, in complex with DARPin binders, in $P2_12_12_1$ [16]. Crystallization in space group $P2_12_12_1$ was possible due to a crystal contact mediated by a DARPin binding specifically to AcrB. Two DARPins were bound to the AcrB trimer, underlining the asymmetry of the molecule.

In AcrB crystals in space groups P1, C2 or R32, crystal contacts are formed between the periplasmic domains or between the cytoplasmic loops connecting the transmembrane helices. This results in a layer-like crystal packing. In many membrane protein crystals, the lattice is built up in layers with weak interactions in one dimension, leading to anisotropic diffraction data and high mosaicity, although these problems did not occur with AcrB crystals. In contrast, the DARPin-mediated crystal lattice provides equal contacts in all three dimensions (Figure 4.1 B). The resulting resolution of 2.5 Å is the highest so far reported for AcrB, which is a consequence of the improved crystal packing mediated by DARPin binders.

The asymmetric structures provided insight into the transport mode, representing a snapshot of three conformations adopted by the protomers during the transport cycle: the access (loose) conformation, binding (tight) conformation and the extrusion (open) conformation. Possible drug extrusion pathways were identified [16] and a rotary pump mechanism was suggested [16,23,24], a hypothesis that was later confirmed by cross-linking experiments [25].

Here, we describe the in-depth analysis of the pool of DARPin binders specific for AcrB selected *in vitro* by ribosome display. When mapping the binding sites on AcrB, we found that all DARPins bind to an overlapping epitope. Comparing three crystal structures of AcrB in complex with different DARPin binders allowed us to study the plasticity of the protein-protein interaction and to evaluate the effectiveness of those DARPins as crystallization chaperones. We also compared affinities, determined the stoichiometry of the complexes and investigated the specificity for AcrB.

4.3 Results

4.3.1 Characterization of DARPin binders

Epitope mapping

Our study aims at the characterization of DARPin binders to AcrB after ribosome display selection [16]. We used ELISA techniques to map the binding sites of various DARPins on AcrB. In order to identify alternative binders from the pool after selection round 3 and 4, 198 clones of each round were tested, comparing the signal representing the interaction of the DARPin with biotinylated AcrB in absence and presence of competing AcrB or the previously described binder 1108_19 [16]. In the following, the latter will be referred to as DARPin#1. We identified 23 specific binders from round 3 and 69 specific binders from round 4, corresponding to a three-fold enrichment. However, all DARPins were competed by DARPin#1. We conclude that they bind to an overlapping epitope on AcrB.

Eight DARPin binders from the fourth selection round, in the following referred to as DARPin#2 to DARPin#9, were chosen for further characterization. Figure 4.2 shows the competition ELISA results for these binders. The signal resulting from the interaction of AcrB with DARPin#1 is reduced to background level when competing with any of the binders DARPin#1-#9, while the non-specific DARPin E3_5 [5] does not have an effect on the signal (Figure 4.2 B). All binders therefore interact with an overlapping epitope. In the reverse experiment, we immobilized DARPin#1-#9 and used DARPin#1 as competitor (Figure 4.2 C). Also here, we observed a significant reduction of the signals. The signal from AcrB binding to DARPin#2 is not reduced as efficiently due to the high affinity of this binder.

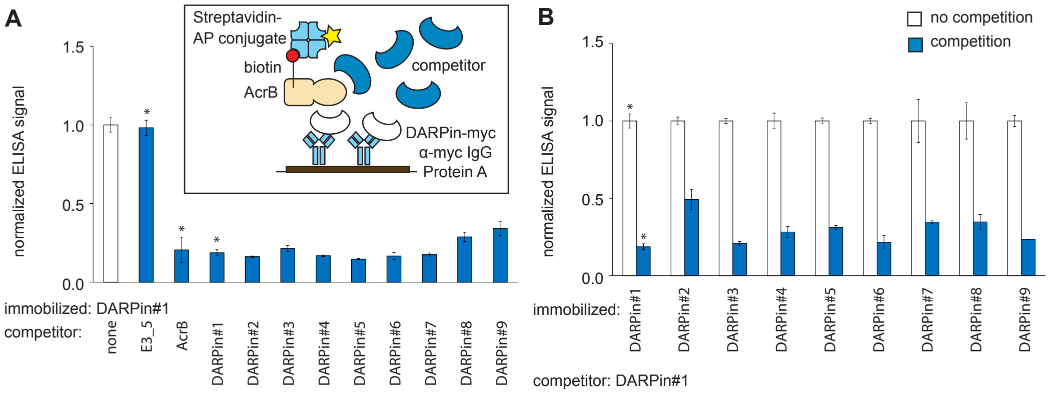


Figure 4.2: Competition ELISA to map binding sites of DARPins on AcrB: (A) Experimental setup: DARPins fused to a C-terminal myc-tag were immobilized via α-myc IgG. Binding of biotinylated AcrB was detected with a streptavidin-alkaline phosphatase conjugate. Competitor (untagged DARPin or unbiotinylated AcrB) was used at a concentration of 1 mM (20-fold excess). (B) Competition of AcrB binding to DARPin#1. (A) and (B) The ELISA signal without competition (white bar) is compared to the signal in presence of a competitor (blue bars). Controls marked with asterisk include competition with the non-specific DARPin E3_5, and competition with AcrB or DARPin#1.

Sequences: Conserved features of DARPin binders

Figure 4.3 shows an alignment of the amino acid sequences with the N3C sequence. DARPins #2-#8 differ from DARPin#1 in 15 to 18 positions. Despite the considerable sequence variation, these binders share several common features with DARPin#1. In the first internal repeat of DARPins #2-#8, we find predominantly hydrophobic side chains such as valine or isoleucine in positions 45 and 46, but also aromatic residues. These AcrB binders all contain a tryptophan in position 48, suggesting an important role for the formation of the binding interface. In positions 56 and 57 of DARPins #2-#8, almost exclusively aromatic amino acids have been selected for. In the second internal repeat, a serine or threonine followed by leucine is found in almost all cases in the β-turn region. The β-turn region of the third internal repeat frequently contains asparagine or histidine in positions 111 or 112. Arginines are found in positions 122 or 123 of DARPins #1, #2, #5-#8. We conclude that DARPins #2-#8 bind to the same epitope as DARPin#1.

In contrast, DARPin#9 contains basic residues in randomized positions 48, 56 and 57 in place of aromatic residues. In fact, this binder differs from DARPin#1 in all randomized positions and also contains framework mutations. An interesting feature is a cysteine at position 122, which was not intended in the library design and musthave evolved during PCR amplification between the selection cycles. Although we propose that the binding site overlaps based on competition ELISA studies, we suggest that DARPin#9 binds AcrB in a different orientation because it does not share any of the features that determine DARPin#1-like binding to AcrB.

In DARPins #6-#8, an interesting framework mutation occurred: the proline initiating helix 1 in the second internal repeat is replaced by leucine. We expect this mutation to have an effect on the relative position of the helices [26].

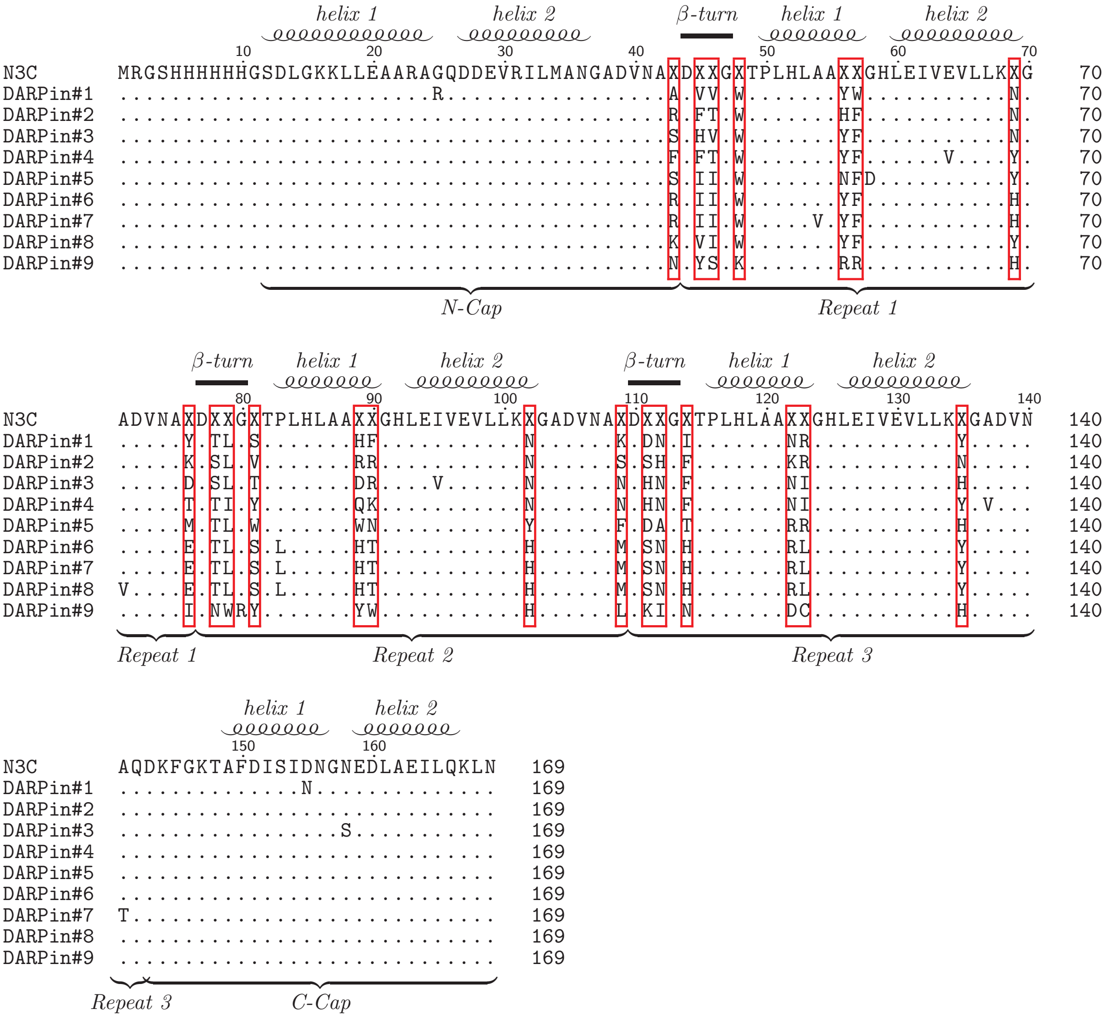


Figure 4.3: Alignment of amino acid sequences of DARPin binders to AcrB. Secondary structure elements are indicated on top. Randomized positions are in a red frame. The alignment was prepared using Texshade [27].

4.3.2 Structures of AcrB-DARPin complexes

Overall binding mode

In addition to the previously known structure of AcrB with DARPin#1 (PDB ID 2J8S [16]), we solved the structures of AcrB in complex with DARPin#2 and DARPin#3 at a resolution of 2.70 Å or 3.34 Å, respectively. The AcrB-DARPin complexes crystallized in 8-10 % PEG4000, 50 mM ADA pH 6.5 and 200 mM $(NH_4)_2SO_4$. This condition had previously been optimized for the AcrB-DARPin#1 complex. Other AcrB-DARPin complexes did not crystallize in this condition. Statistics of processing and refinement are summarized in Table 4.1. The data could be processed in the same space group, $P2_12_12_1$, with similar unit cell parameters. However, in the AcrB-DARPin#3 structure, the c-axis is more than 10 Å longer than in the other two structures.

Two DARPins are bound to the AcrB trimer in all cases (Figure 4.4 A). The binders interact with the DC subdomain which is part of the TolC docking domain of AcrB, located at the tip of the periplasmic part of the transporter. The DC subdomain contains a 4-stranded beta-sheet and the epitope is mainly formed by strand 802-811 on the surface of AcrB.

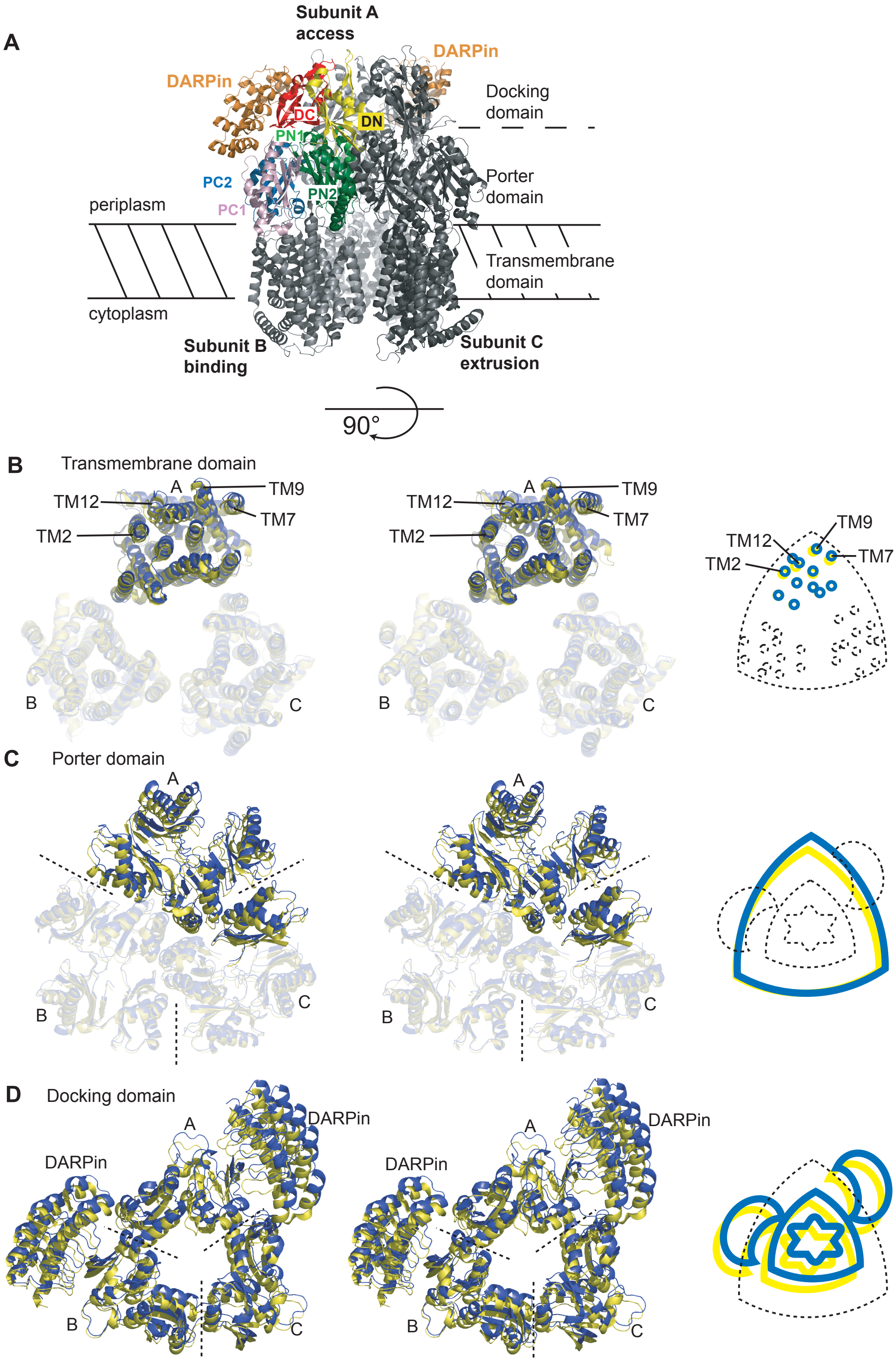
Table 4.1: Data collection and refinement statistics.

Data collection	AcrB-DARPin#1	AcrB-DARPin#2	AcrB-DARPin#3
PDB ID	2J8S [16]	3NOC	3NOG
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a = 146.2	a = 146.0	a = 145.4
	b = 157.4	b = 158.7	b = 158.0
	c = 246.0	c = 244.5	c = 258.6
Resolution (Å)	35.00-2.54	50.00-2.70	50.00-3.34
(highest shell)	(2.70-2.54)	(3.00-2.70)	(3.60-3.34)
No. of unique reflections	177557	154197	86818
Redundancy	9.1	4.5	6.6
Completeness (%)	95.0 (79.0)	99.6 (99.8)	99.8 (99.7)
R_{merge}	0.081 (0.321)	0.116 (0.365)	0.128 (0.363)
$I/\sigma(I)$	20.75 (6.1)	13.71 (4.02)	14.98 (3.9)

Refinement			
Resolution (Å)	25.00-2.54	50.00-2.70	50.00-3.34
Number of reflections	177557	154197	86818
R (R_{free})	22.9 (27.0)	24.2 (26.8)	25.7 (30.7)
No. of waters	350	438	0
Mean B factors (^{2})			
AcrB	42.3	60.4	76.1
DARPin (chain D)	42.2	66.1	89.5
DARPin (chain E)	54.2	101.2	161.0
RMS deviations			
- bond lengths (Å)	0.008	0.008	0.014
- bond angles (degrees)	1.3	0.8	1.5
Ramachandran statistics			
- allowed (No. of residues)	92.04 % (3144)	97.2 % (3279)	93.96 % (3160)
- generously allowed	5.59 % (191)	2.79 % (94)	5.86 % (197)
- disallowed	2.37 % (81)	0.03 % (1)	0.18 % (6)

When comparing the overall structure with the AcrB-DARPin#1 complex, the RMSD value is 0.490 ± 0.289 Å for the AcrB-DARPin#2 complex and 1.540 ± 0.571 Å for the AcrB-DARPin#3 complex. The crystal-contact mediating DARPin, bound to subunit B of AcrB, is in the same position as in the other two structures, but the DARPin bound to subunit A is shifted due to a change in the conformation of AcrB. This becomes evident when we superimpose the transmembrane region of the AcrB-DARPin#3 complex with the corresponding region in AcrB-DARPin#1. A small outward shift of transmembrane helices 2, 7, 9 and 12 is observed in subunit A (Figure 4.4B). Since these transmembrane helices connect to the periplasmic part of the molecule, this causes a rigid-body outward movement of the periplasmic domains of this subunit and the neighboring chain C, explaining the substantial RMS deviation (Figure 4.4C,D). Subunit A is involved in a crystal contact with subunit C via their periplasmic domains (Figure 4.5), which leads to the 10 Å difference in the length of one unit cell edge described above (Table 4.1). This difference in the conformation of subunit A of AcrB in complex with DARPin#3 represents a subconformation of the access form of the protomer and illustrates the potential of AcrB to adopt different conformations necessary for its function.

Figure 4.4: Comparison of crystal structures of AcrB-DARPin complexes. (A) AcrB-DARPin structure overview: AcrB subunits (gray) and DARPin binders (orange) are shown in cartoon representation viewed along the membrane plane. In subunit B, subdomains PN1 (light green), PN2 (dark green), PC1 (rose), PC2 (blue), DN (yellow) and DC (red) are labeled in the respective color. (B), (C), (D) Differences in the three DARPin AcrB complex. The transmembrane domain of AcrB-DARPin#3 was superimposed with the corresponding region in AcrB-DARPin#1. The transmembrane domain (B), the porter domain (C) and the docking domain with bound DARPins (D) are shown in stereo viewed from the periplasmic side (90ž rotation of (A) as indicated). The yellow molecule represents the AcrB-DARPin#1 complex and AcrB-DARPin#3 is shown in blue. Transmembrane helices TM2, TM7, TM9 and TM12 that are affected by the conformational rearrangement in the AcrB-DARPin#3 crystal are labeled.


A
Subunit A
access
DARPin
DARPin
Docking
domain
DC
DN
PN1
PC2
PN2
PC1
Porter
domain
periplasm
Transmembrane
domain
cytoplasm
Subunit B
binding
Subunit C
extrusion
90°
B
Transmembrane domain
A
TM9
TM12
TM7
TM2
B
C
A
TM9
TM12
TM7
TM2
B
C
TM9
TM12
TM7
TM2
C
Porter domain
A
B
C
A
B
C
D
Docking domain
DARPin
A
DARPin
B
C
DARPin
A
DARPin
B
C

Resolution resulting from crystal packing

Even though the molecules are arranged in a very similar manner in all three structures of the complex, the resolution that could be obtained is quite different comparing the three DARPins: AcrB in complex with DARPin#1 was refined to 2.54 Å, with DARPin#2 to 2.70 Å, and the AcrB-DARPin#3 complex could only be refined to 3.34 Å. A closer look at the crystal contact reveals that this DARPin cannot form the crystal contact equally well as DARPin#1 or #2 (Table 4.2, Figure 4.6). A slight difference in angle reduces the number of hydrogen bonds to 4 instead of 7 and the buried surface area from 444.4 Å^2 to 409.8 Å^2. This is a consequence of the different conformation of AcrB in complex with DARPin#3 as described above, resulting in modified cell parameters.

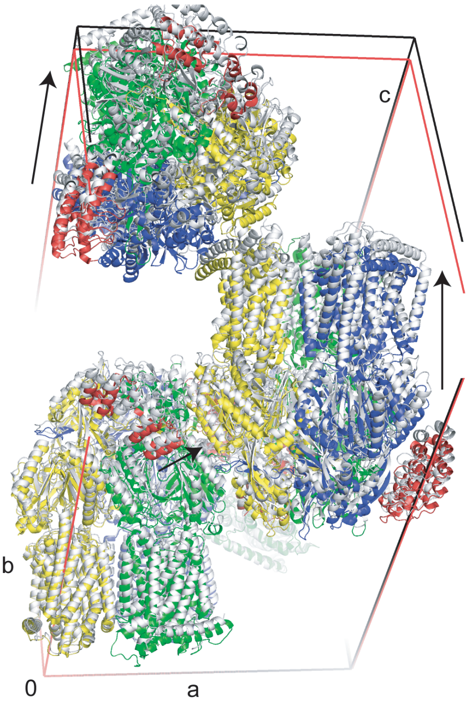


Figure 4.5: The unit cells of AcrB-DARPin crystals are drawn with a common origin. The unit cell content of AcrB-DARPin#1 crystals shown in colors (chain A, green; chain B, blue; chain C, yellow; chain D and chain E (DARPins), red; unit cell edges, red) is compared with the packing of AcrB-DARPin#3 (white, unit cell edges, black). The outward movement of the periplasmic part of chain A in the AcrB-DARPin#3 complex affects the crystal contact with chain C and causes a shift of the symmetry-related molecules indicated by black arrows. As a consequence, the c-edge is 12 Ålonger in crystals of AcrB-DARPin#3 as compared to AcrB-DARPin#1.

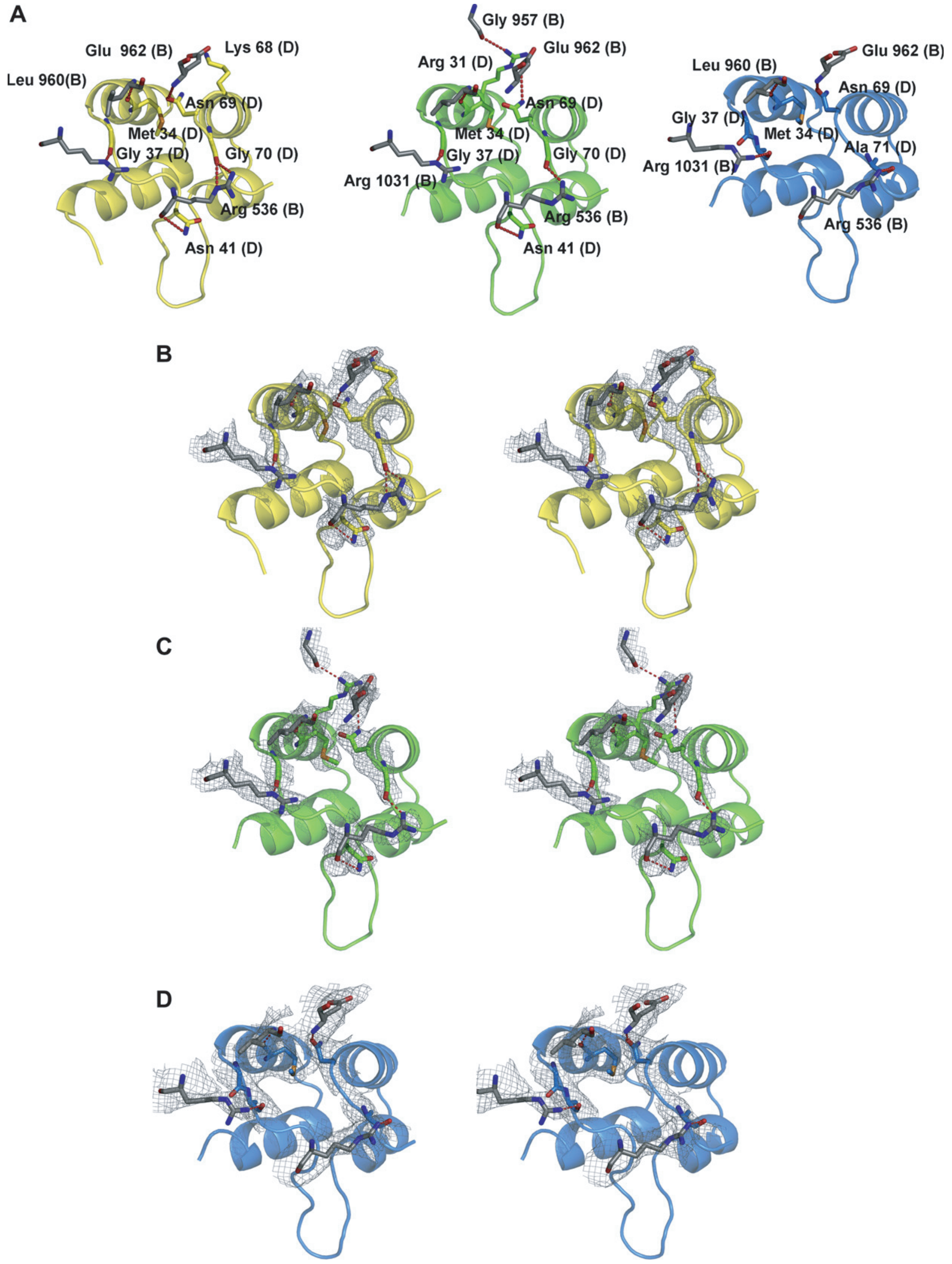


Figure 4.6: Close-up on the DARPin mediated crystal contact (DARPin#1, yellow; DARPin#2, green; DARPin#3, blue; AcrB residues, gray). (A) The DARPin viewed from the symmetry mate is drawn in cartoon representation. Residues involved in a hydrogen bond are shown as sticks. Hydrogen bonds are represented by red dashed lines. Stereo views of the close-up on DARPin#1 (B), DARPin#2 (C), DARPin#3 (D) with a 2Fo-Fc electron density map contoured at 1 sigma around residues involved in a hydrogen bond depicted as a gray mesh. The 2Fo-Fc maps were calculated in PHENIX [28].

Table 4.2: Hydrogen bonds or *salt bridges* in the DARPin-mediated crystal contact with the cytoplasmic loops of AcrB subunit B.

AcrB	DARPin#1	(Å)	DARPin#2	(Å)	DARPin#3	(Å)
Arg 536 (O)	Asn 41 (ND2)	3.3	Asn 41 (ND2)	3.7		
Arg 536 (NE)	Gly 70 (O)	3.0	Gly 70 (O)	2.6		
Arg 536 (NH2	Gly 70 (O)	2.8			Ala 71 (O)	3.4
Gly 957 (O)			Arg 31 (NH1)	3.4		
Leu 960 (N)	Met 34 (O)	3.1	Met 34 (O)	2.8	Met 34 (O)	3.0
Glu 962 (OE2)			Asn 69 (ND2)	3.8		
Glu 962 (N)	Asn 69 (OD1)	2.9	Asn 69 (OD1)	3.6	Asn 69 (OD1)	2.8
Glu 962 (OE1, OE2)	*Lys 68 (NZ)*	*3.7*				

AcrB-DARPin interface

The three structures of AcrB in complex with three different DARPins allow the direct comparison of the interaction interface.

The buried surface area of the crystal-contact mediating DARPin (chain D) with AcrB is relatively large with 1163.7 Å^{2}2 for the AcrB-DARPin#1 complex, 903.3 Å^{2} for the AcrB-DARPin#2 complex and 1051.7 Å^{2} for the AcrB-DARPin#3 complex. We have performed a computational alanine scan [29] to identify residues that are important for DARPin binding to AcrB. On the AcrB surface, tryptophan 809 and tyrosine 811 were found to be the only residues which are crucial for the interaction. These residues form a hydrophobic core with the randomized positions of the first internal repeat. Key residues on the DARPin side are tryptophan 48, tryptophan or phenylalanine 57, and a small side chain in position 46 that can be valine, isoleucine or threonine. In addition, framework residues contribute to the hydrophobic interactions at positions 53, 77, and 145. The variation in the randomized positions leads to differences in the DARPin interaction surface, resulting in a significantly lower shape complementarity of 0.477 for DARPin#3 as compared to 0.643 for DARPin#1 or 0.684 for DARPin#2. The latter two values are in the range typically observed for antibody-antigen interactions [30].

The polar interactions between AcrB and the respective DARPin are summarized in Figure 4.7 (see also Tables 4.3/4.4). Equivalent hydrogen bonds in the core of the interface are frequently observed in all three structures: For example, tryptophan 809 forms a hydrogen bond via the hydroxyl group of serine or threonine 78. Also the polar interaction of asparagine or histidine in position 112 (in DARPin#3 also histidine 111) is well-conserved. Salt bridges seem to be quite exchangeable.

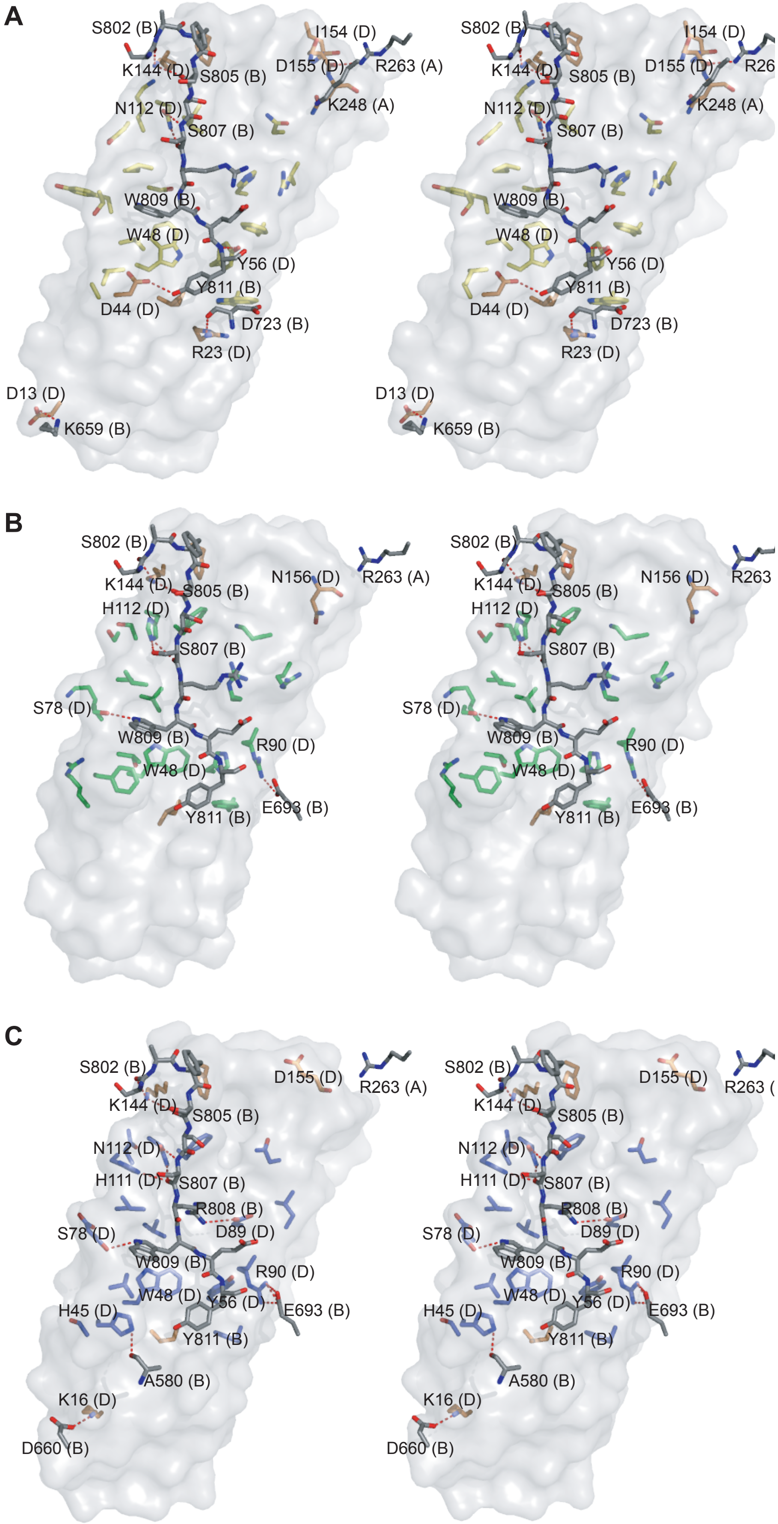


Figure 4.7: Stereo view of the AcrB-DARPin binding interface. The DARPins surface is shown in light gray. Residues involved in the interaction are drawn as sticks. DARPin residues in randomized positions are colored yellow (A, DARPin#1), green (B, DARPin#2) or blue (C, DARPin#3), and framework residues are orange. Relevant AcrB residues are shown in gray.

Table 4.3: Hydrogen bonding (*salt bridges*) between chain D of DARPin#1-#3 and AcrB.

AcrB	DARPin#1	(Å)	DARPin#2	(Å)	DARPin#3	(Å)
Subunit A						
Lys 248 (NZ)	Asn 156 (OD1)	3.0				
Arg 263 (NH2)	Asn 155 (O)	2.7	Asn 156 (O)	3.4	Asp 155 (O)	3.8
Subunit B						
Ala 580 (O)					His 45 (NE2)	3.1
Lys 659 (NZ)	*Asp 13 (OD2)*	*3.23*				
Asp 660 (OD1)					*Lys 16 (NZ)*	*2.4*
Glu 693 (OE2)			*Arg 90 (NH1)*	*3.3*	*Arg 90 (NH1)*	*2.6*
Asp 723 (O)	Arg 23 (NH1)	3.2				
Glu 734 (OE2)	*Lys 147 (NZ)*	*3.7*	*Lys 147 (NZ)*	*3.6*	*Lys 147 (NZ)*	*4.0*
Ser 802 (O)	Lys 144 (NZ)	2.6	Lys 144 (NZ)	3.1	Lys 144 (NZ)	3.1
Ser 805 (O)	Lys 144 (NZ)	3.0	Lys 144 (NZ)	2.4	Lys 144 (NZ)	3.4
Ser 807 (O)	Asn 112 (ND2)	2.7	His 112 (NE2)	3.2	His 111 (NE2)	3.2
Ser 807 (O)					Asn 112 (ND2)	2.5
Ser 807 (N)	Asn 112 (OD1)	3.3			Asn 112 (OD1)	3.2
Ser 807 (OH)			His 112 (NE2)	2.83		
Arg 808 (NH1)					*Asp89 (OD2)*	*3.4*
Trp 809 (NE1)			Ser 78 (OH)	2.9	Ser 78 (OH)	3.2
Tyr 811 (OH)	Asp 44 (OD2)	3.1				
Tyr 811 (N)	Tyr 56 (OH)	3.0			Tyr 56 (OH)	3.6
Buried surface area (Å^2)		1163.6		855.8		1037.5
Shape complementarity		0.643		0.684		0.477

Table 4.4: Hydrogen bonding (*salt bridges*) between chain E of DARPin#1-#3 and AcrB.

AcrB	DARPin#1	(Å)	DARPin#2	(Å)	DARPin#3	(Å)
Subunit C						
Arg 259 (NZ)					Asp 155 (O)	3.8
Lys 248 (NZ)			Asn 156 (O)	3.0		
Arg 263 (NH1)	Asn 155 (O)	2.6			Asn 156 (O)	3.6
Arg 263 (NH2)	Ile 154 (O)	3.2	Ile 154 (O)	3.8	Asn 156 (O)	3.4
Subunit A						
Glu 693 (OE2)			*Arg 90 (NH2)*	*2.7*		
Asp 723 (O)	Arg 23 (NH1)	3.1	Arg 23 (NH1)	3.3		
Glu 734 (OE2)	*Lys 147 (NZ)*	*2.7*				
Ser 802 (O)	Lys 144 (NZ)	2.6	Lys 144 (NZ)	2.4		
Ser 805 (O)	Lys 144 (NZ)	3.0	Lys 144 (NZ)	2.7		
Ser 807 (O)	Asn 112 (ND2)	2.6	His 112 (NE2)	3.1	Asn 112 (ND2)	3.2
Trp 809 (NE1)	Thr 78 (OG1)	2.9	Ser 78 (OH)	3.1	Ser 78 (OH)	2.9
Tyr 811 (N)	Tyr 56 (OH)	3.0				
Buried surface area (Å^2)		993.0		901.3		678.4
Shape complementarity		0.655		0.591		0.385

In order to understand what makes this epitope so attractive for DARPin binding, we looked at the surface of AcrB (Figure 4.8). Tryptophan 809 and tyrosine 811 are well-accessible for protein-protein interaction, providing a hydrophobic patch and as well as a hydrogen bond acceptor. The periplasmic part of the transporter does not contain other similarly exposed tryptophan residues. A few tryptophans are found at the cytoplasmic end of the transmembrane helices. They are probably shielded by the lipid bilayer or the detergent micelle. The presented structures show DARPin binding to subunit A (access/loose) and subunit B (binding/tight), but not to subunit C (extrusion/open) despite the presence of the same hot-spot binding site. However, DARPin binding is not compatible with the conformational rearrangements in this protomer because of a steric clash with a helical portion of subdomain PC2 (Figure 4.9).

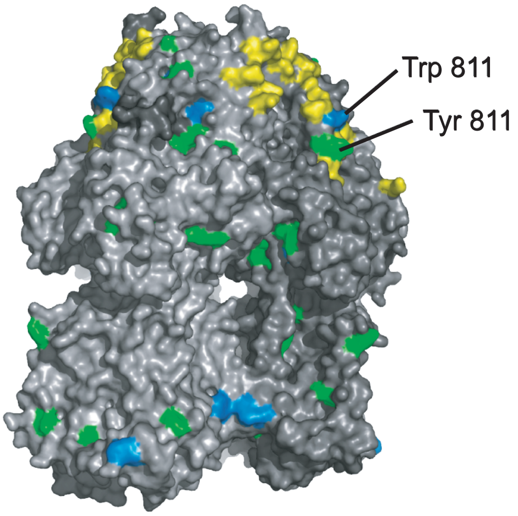
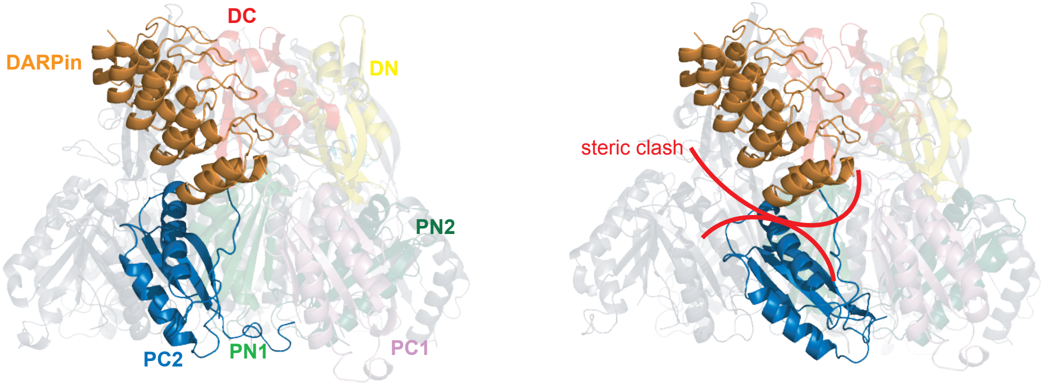


Figure 4.8: Surface representation of AcrB in different shades of gray. Residues interacting with DARPin binder DARPin#1 are colored in yellow. All tryptophan residues are shown in blue tyrosines are cyan. The binding site for DARPin#1 contains a well-accessible tryptophan with a neighboring tyrosine.



binding / tight conformation

extrusion / open conformation

Figure 4.9: Cartoon representation of the periplasmic part of AcrB in the binding / tight conformation (left) or the extrusion / open conformation (right) in the same orientation. The subunits are labeled in the respective color. The DARPin is depicted in orange. A steric clash with PC2 subdomain prevents DARPin binding to the extrusion / open conformation of the AcrB protomer.

4.3.3 Biophysical characterization of the DARPin-AcrB interaction

Affinity

The affinity of DARPins #1-#9 to their target was determined using Surface Plasmon Resonance (BIAcore). The kinetic parameters obtained are summarized in Table 4.5. All dissociation constants are in the medium to low nanomolar range with values between 1.5 and 89 nM. Binding is characterized by relatively fast off-rates. DARPin#2 shows the highest affinity to AcrB, combining the fastest on-rate with the slowest off-rate. Despite a less optimal shape complementarity of the AcrB-DARPin#3 interaction, this binder displays a high affinity to AcrB with a dissociation constant of 22.4 nM. The dissociation constants of DARPin#6-#8 differ significantly with values of 7.6, 61.3 and 89.3 nM, even though the sequence of these DARPins only deviates in single framework residues (Figure 4.3). The randomized positions are identical. This exemplifies the importance of the exact spatial arrangement of the randomized patch by the scaffold structure. We conclude that these substitutions destabilize the interaction and might have been introduced at a later stage of the selection process.

Table 4.5: Kinetic parameters of DARPin binding to AcrB determined by surface plasmon resonance.

DARPin	$\mathrm{k_{on}}$ $[\mathrm{M^{-1}s^{-1}}]$	$\mathrm{k_{off}}$ $[\mathrm{s^{-1}}]$	$\mathrm{K_D}$ [nM]	$\mathrm{R_{max}}$ [RU]	$\mathrm{Chi^2}$ $[\mathrm{RU^2}]$
DARPin#1	$8.7 \cdot 10^5$	$2.4 \cdot 10^{-2}$	28.0		
DARPin#2	$6.0 \cdot 10^6$	$9.0 \cdot 10^{-3}$	1.5	25.6	0.441
DARPin#3	$6.5 \cdot 10^5$	$1.5 \cdot 10^{-2}$	22.4	25.2	0.350
DARPin#4	$2.0 \cdot 10^6$	$3.7 \cdot 10^{-2}$	18.7	22.7	0.301
DARPin#5	$2.0 \cdot 10^6$	$6.4 \cdot 10^{-2}$	3.2	24.9	0.206
DARPin#6	$3.7 \cdot 10^6$	$2.8 \cdot 10^{-2}$	7.6	26.6	0.214
DARPin#7	$1.9 \cdot 10^5$	$1.2 \cdot 10^{-2}$	61.4	18.6	0.247
DARPin#8	$1.6 \cdot 10^5$	$1.4 \cdot 10^{-2}$	89.3	14.0	0.426

Stoichiometry

The previously published crystal structure of the AcrB-DARPin complex [16] shows an asymmetric conformation of AcrB with DARPins binding to two of the three chains of the trimer. We investigated the stoichiometry of AcrB in complex with the other binders in solution. To this end, we used gel filtration to isolate the complex and quantified each component using a protein capillary electrophoresis chip.

Not all DARPins bind AcrB with the same stoichiometry (Table 4.6). In agreement with the crystal structures, two molecules of DARPin#1, #2 or #3 form a complex with the AcrB trimer. In contrast, we find one binder for DARPins #4, #5 and #9, while DARPin#6 and #7 form a complex composed of three DARPins per AcrB trimer. Only AcrB-DARPin complexes with two binders could be crystallized in the same crystal form which requires the crystal-contact mediating DARPin to be bound to subunit B (binding/tight conformation) of AcrB. We conclude that only this stoichiometry is compatible with lattice formation in this particular crystal form. If only one binder is present, it will likely interact with subunit C, which is in a conformation that differs the most in the asymmetric structures determined so far. Binders to all three subunits of AcrB might stabilize a more symmetric conformation of the transporter. This analysis again illustrates the potential of ribosome display selection of binders, allowing to obtain information regarding different functional states of a target protein.

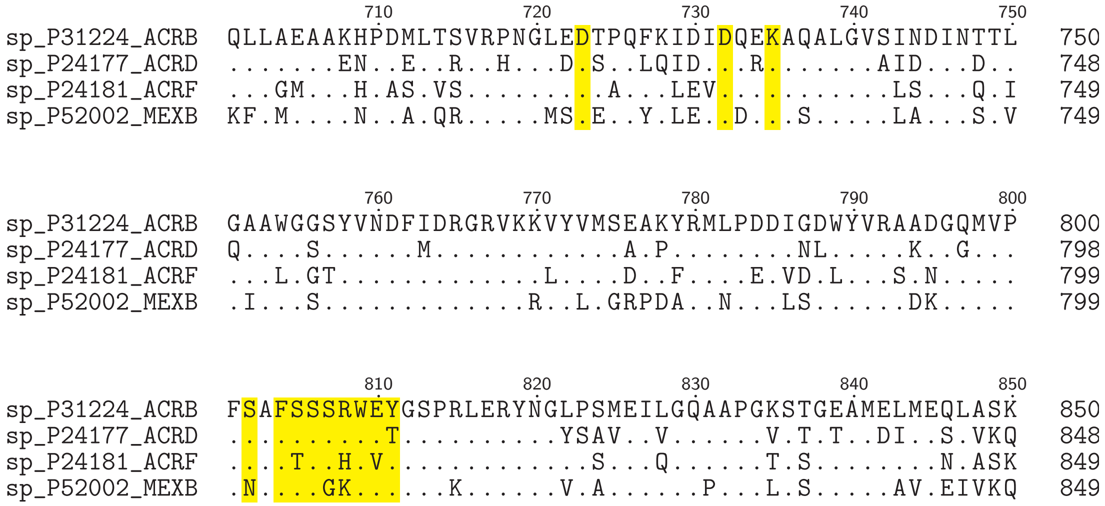
Table 4.6: Specificity of DARPin binders for AcrB.

DARPin	Stoichiometry (DARPins/AcrB trimer)	Specificity for AcrB			
		AcrB	AcrD	AcrF	MexB
DARPin#1	2	+++	-	-	-
DARPin#2	2	+++	-	+	-
DARPin#3	2	+++	-	+	-
DARPin#4	1	+++	-	-	-
DARPin#5	1	+++	-	+	-
DARPin#6	3	+++	-	+	-
DARPin#7	3	+++	-	-	-
DARPin#8	n.d.	+++	-	-	-
DARPin#9	1	+++	-	-	-
controls					
α-AcrF		-	-	+++	-
α-MexB		-	-	-	+++

Specificity

We further investigated the binding of DARPins to proteins homologous to AcrB, namely AcrD (identity 66 %, similarity 80 %) and AcrF (identity 77 %, similarity 88 %) from *Escherichia coli* as well as MexB from *Pseudomonas aeruginosa* (identity 69 %, similarity 83 %). The crystal structure of AcrB in complex with DARPin#1 revealed a number of residues conserved in all four RND transporters (Figure 4.10) that interact with the bound DARPin. The results of the specificity analysis are summarized in Table 4.5. First, we used a streptavidin alkaline phosphatase conjugate to detect binding of chemically biotinylated RND transporters to immobilized DARPins by ELISA. Binding of DARPin#2, #3, #5 and #6 to AcrF, but not AcrD or MexB, could be detected (Figure 4.11A), while DARPin#1, #4, #7, #8 and #9 were specific for AcrB. We used DARPins raised against AcrF and MexB as a control (Sennhauser, unpublished). These binders were very specific for their respective target.

In order to compare the background interaction to binding by AcrB, we compared the surface plasmon resonance signal (Figure 4.11B). The results agree well as background binding to AcrF was observed for the listed DARPins. The cross-specific binders are among the DARPins that show the highest affinity for AcrB. However, the signal from interaction with AcrF is



```
                        710       720       730       740       750
sp_P31224_ACRB  QLLAEAAKHPDMLTSVRPNGLEDTPQFKIDIDQEKAQALGVSINDINTTL  750
sp_P24177_ACRD  .......EN..E..R..H...D.S..LQID...R.......AID...D..  748
sp_P24181_ACRF  ...GM...H.AS.VS.........A...LEV...........LS...Q.I  749
sp_P52002_MEXB  KF.M....N..A.QR.....MS.E..Y.LE..D...S.....LA...S.V  749

                        760       770       780       790       800
sp_P31224_ACRB  GAAWGGSYVNDFIDRGRVKKVYVMSEAKYRMLPDDIGDWYVRAADGQMVP  800
sp_P24177_ACRD  Q....S......M............A.P........NL.....K..G...  798
sp_P24181_ACRF  ...L.GT.............L....D..F....E.VD.L...S.N.....  799
sp_P52002_MEXB  .I...S..............R..L.GRPDA..N...LS......DK.....  799

                        810       820       830       840       850
sp_P31224_ACRB  FSAFSSSRWEYGSPRLERYNGLPSMEILGQAAPGKSTGEAMELMEQLASK  850
sp_P24177_ACRD  ..........T..........YSAV..V......V.T.T..DI..S.VKQ  848
sp_P24181_ACRF  ....T..H.V..............S...Q......T.S........N.ASK  849
sp_P52002_MEXB  .N....GK......K......V.A......P...L.S.....AV.EIVKQ  849
```

Figure 4.10: Alignment of RND transporters AcrD, AcrF and MexB with AcrB. Residues of AcrB interacting with DARPin#1 and their corresponding residues in the other proteins are highlighted in yellow. The alignment was prepared using Texshade [27]

much lower than for the interaction of DARPin DARPins with AcrB suggesting much lower affinity constants. In agreement with the ELISA study, no interaction could be seen between DARPin DARPins and AcrD or MexB. Interestingly, AcrF is less conserved (3 substitutions) in the residues that are involved in binding to the DARPin in the DARPin#1-AcrB complex than AcrD (only 1 substitution) (Figure 4.10). However, in terms of overall homology, AcrF is closest to AcrB of all homologues tested, which seems to cause cross-reactivity with AcrB binders.

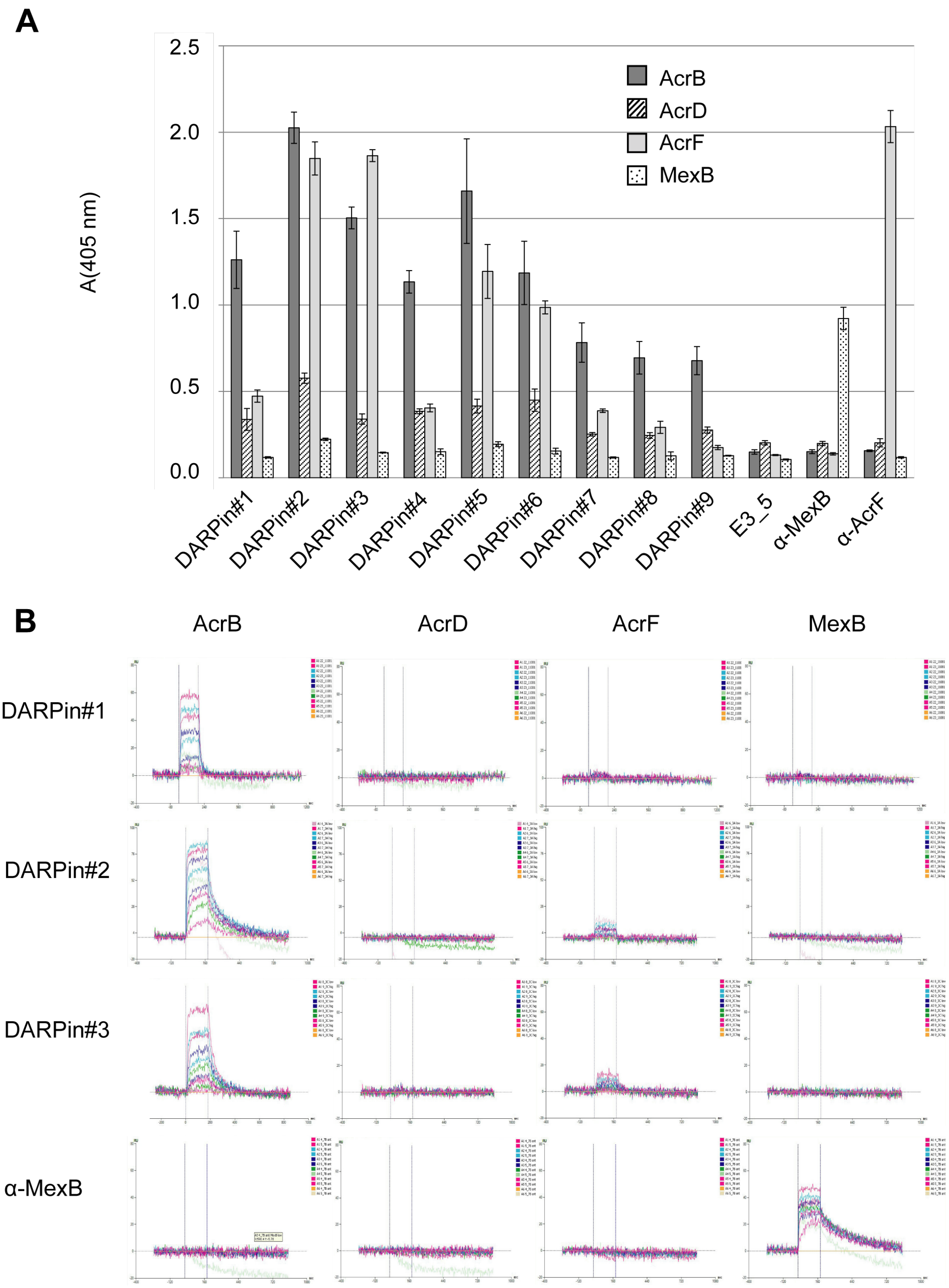


Figure 4.11: Specificity of DARPins for AcrB: (A) ELISA detecting binding of DARPins to AcrB, AcrD, AcrF or MexB (B) Surface plasmon resonance signals of DARPins #1-#3 and α-MexB with AcrB, AcrD, AcrF and MexB.

4.4 Discussion

4.4.1 Molecular recognition and evolutionary protein design

The generation of protein binders by various methods of evolutionary protein design is a powerful tool in modern molecular biology. In addition to antibodies and fragments thereof in diverse formats, alternative scaffolds have been developed to overcome the limitations of antibodies. In order to improve libraries for binder selection, it is important to understand the factors governing molecular recognition. Likewise, the study of synthetic binding interfaces helps us understand both natural and designed protein-protein interactions [31]. In this publication, we present the first structural analysis of three DARPins binding to the same epitope. In addition to the crystal structures of AcrB-DARPin complexes which allow the detailed analysis of the plasticity of the interaction, the DARPins were biophysically characterized.

The analysis of a phage display library of human growth hormone (hGH) binding to its receptor is the most comprehensive mutational analysis of a protein-protein interface to date [32]. Analyzing a sequence database, the authors found that most residues are tolerant to substitution, and only few residues in a hot-spot area are essential for binding. Rather than probing the adaptability of a naturally evolved interface, our study analyzes a pool of designed proteins selected for AcrB binding. The size difference suggests that there are many possibilities for the small DARPins to interact with the target protein. Further experiments were therefore necessary to distinguish between sequence diversity resulting from alternative binding sites or low selection pressure on certain positions that are not involved in binding.

4.4.2 Dominant epitope

Our pool of DARPin binders from a ribosome display selection contains predominantly binders to overlapping epitopes on AcrB. The binding site is mainly formed by one sheet of a beta strand containing three serines followed by a tyrosine. These side chains are overrepresented in the complementarity determining regions of antibodies [33]. Phage display libraries with a minimalist amino acid composition, allowing only tyrosine, alanine, aspartate and serine in the paratope, have been successfully used to select high-affinity binders to protein targets [34]. Further limitation to tyrosine and serine showed that a binary code, with only these two residues forming the interaction surface, is sufficient for specific molecular recognition, both in the case of antibody-antigen interaction [35] and protein binding by a FN3 scaffold [36].

In addition to the serines and tyrosine, a neighboring tryptophan contributes to the formation of a surface area that is apparently more attractive for protein-protein interactions than others. The question arises whether or not there is a functional meaning to the existence of such a hydrophobic patch on the periplasmic domain of AcrB. Although the current model of the AcrAB-TolC complex [37] suggests that the DARPin binding site would not be involved in the formation of the functional assembly [38], we propose that this area is part of the interaction surface of AcrB and AcrA. However, further experimental evidence is needed to clarify this issue.

If a more diverse pool is desired, performing selections to the target molecule in complex with a dominant binder might be helpful. By this means, the shielding of the most attractive epitope would allow the enrichment of alternative binders.

4.4.3 Plasticity of the interface

Structural information on three of the AcrB-DARPin complexes binding to the same epitope puts us in the position to identify the residues crucial for binding. This information can then be used to recognize redundant binders on the sequence level. In agreement with the hGH study, we find that only conservative changes occur around the hydrophobic core of the DARPin-AcrB interface. Intermolecular polar interactions certainly stabilize the interaction, but seem to be rather exchangeable. The highest affinity DARPin forms fewer hydrogen bonds and salt bridges than the other two binders, but exhibits the highest shape complementarity.

4.4.4 DARPins for the improvement of membrane protein crystal quality

Our study aimed at evaluating DARPins for the crystallization of membrane proteins. We observe different stoichiometries critically affecting crystallization of the respective AcrB-DARPin complex. Crystals grew when two DARPins formed a complex with the AcrB trimer, although only one of the binders was involved in a crystal contact.

Interestingly, DARPins that bind substoichiometrically still compete for AcrB binding as judged by ELISA. Even when only one or two DARPins bind to the homotrimer, leaving the binding site on the other subunits unoccupied, mixed complexes do not seem to be formed. In case of AcrB, the unoccupied binding sites are fixed in a conformation which is not favorable for the interaction with a different DARPin. This observation supports the hypothesis that synthetic binding proteins can trap a conformation of their target protein. Crystallization chaperones thereby reduce conformational heterogeneity and facilitate nucleation of crystal growth.

Even though the DARPins bind to the same site, the resolution obtained varied substantially. Since the binders can only be selected for optimal interaction with the target protein rather than formation of a crystalline lattice, it is important to consider screening of several DARPins to achieve the best possible diffraction data. Although monodispersity at high concentrations and reasonable affinity of the crystallization chaperone to the target are prerequisites for a successful crystal growth, evaluation of biophysical properties of the complex in solution does not guarantee optimal diffraction properties. Screening of several seemingly redundant binders is therefore worthwhile to obtain the best-possible outcome. Construction of a focused library in which residues determining the binding interface are kept constant would be a comprehensive approach. Here, also optimization of surface residues potentially involved in crystal contacts could be considered, for example by surface entropy reduction [39,40].

4.5 Experimental procedures

4.5.1 Expression and purification

AcrB was expressed from a pET28 vector (Novagen) with a C-terminal *His*-tag and purified as described previously [16] using buffer A (20 mM Tris pH 7.5, 150 mM NaCl, 0.03 % DDM). The same procedure was used for the expression and purification of the homologs AcrD, AcrF and MexB.

For enzymatic biotinylation [41], an expression vector for AcrB carrying a C-terminal *His*- and *avi*-tag was constructed. A DNA fragment encoding the *Avi* tag sequence flanked by the restriction sites NheI and BamHI was formed by annealing the two oligonucleotides avitag_for (5'-CAT GGG CGC TAG CGG TCT GAA CGA TAT CTT CGA AGC TCA GAA AAT CGA ATG GCA CGA AGG ATC CTA AT-3') and avitag_rev (5'- CTA GAT TAG GAT CCT TCG TGC CAT TCG ATT TTC TGA GCT TCG AAG ATA TCG TTC AGA CCG CTA GCG CC -3'). The fragment was ligated into the *E. coli* cloning vector pGEM using the NcoI and XbaI restriction sites, yielding pGEM_Avi. The gene of *acrB* from *Escherichia coli* devoid of NcoI sites (Murakami and van Veen, unpublished) was amplified with the primers acrB_HisC_AviC_for (5'-ATC CAT GGG GCC TAA TTT CTT TAT CGA TCG CCC G-3') and acrB_HisC_AviC_rev (5'-ATG CTA GCG TGA TGG TGA TGG TGA TGG TGA TGC TGC AGA TGA TGA TCG ACA GTA TGG CTG TG-3') and cloned via NcoI/NheI into pGEM_AviC yielding $\mathrm{pGEMAcrB_{AviC}}$. The tagged acrB genes were sub-cloned into pBAD24 using the restriction sites NcoI and XbaI resulting in $\mathrm{pBADAcrB_{AviC}}$, which was used for protein production.

Expression in Escherichia coli strain C43(DE3) was induced with 0.01 % D-arabinose at $\mathrm{OD_{600}}$ = 0.5 for 4 h at 30 degrees. After purification via the C-terminal *His*-tag, AcrB-avi was biotinylated over night at 4 degrees at a concentration of 10 μM with 50 μg of BirA in presence of 5 mM ATP, 10 mM magnesium acetate and 1.2 fold excess of biotin in a total volume of 2500 μl of buffer A supplemented with 10 % glycerol. Excess biotin as well as BirA were removed by gel filtration in buffer A-10 % glycerol on a Superdex 200 10/300 GL column (Pharmacia).

Chemical biotinylation was performed after purification of RND transporters by IMAC using the Tris-free buffers 10 mM, 50 mM and 200 mM imidazole/HCl pH 7.5, 150 mM NaCl, 10 % glycerol and 0.03 % DDM. The protein was biotinylated with 1 mg/ml EZ-Link Sulfo-NHS-LC-biotin (Pierce) for 30 min on ice. Excess biotin was removed by gel filtration in buffer A/10 % glycerol on a Superdex 200 10/300 GL column (Pharmacia).

4.5.2 Ribosome display

All binders used in this study originate from the third and fourth round of a ribosome display selection described previously [16]. Briefly, binders to AcrB were selected from a N3C DARPin library [5,10] after pre-panning with neutravidin, BSA and biotinylated maltose-binding protein (MPB). The panning steps were performed at room temperature using buffer A.

4.5.3 ELISA

Crude extract ELISA with or without competition was performed as described by Huber *et al.* [11] with some modifications, using TBS (50 mM Tris/HCl pH 7.4, 150 mM NaCl), TBS-BSA (TBS with 0.5 % BSA), TBS-D (TBS containing 0.01 % DDM) or TBS-D-BSA (TBS-BSA containing 0.03 % DDM) for binding and washing.

To prepare crude extract in a 96-well format, 1.2 ml per well of 2xYT medium containing 0.5 % glucose and 50 μg/ml ampicillin were inoculated with a single colony harboring the expression plasmid pQE30-DARPin-(myc)$_5$. The cultures were incubated at 37 degrees and 450 rpm. The next morning, 300 μl of the overnight culture was added to 900 μl of fresh 2xYT with 100 μg/ml ampicillin. Expression was induced after 2 h at 37 degrees and 450 rpm with IPTG at a final concentration of 0.25 mM. The cells were harvested after 3 h and lysed using 50 μl B-PERII (Pierce) for 10 min at room temperature. The lysate was diluted in 950 μl TBS and stored at 4 degrees.

Immunoplates (Nunc) were coated with 100 μl per well of 5 ng/μl of Protein A in PBS, pH 7.4, over night at 4 degrees. Each step was followed by washing three times with 300 μl TBS or TBS-D after addition of the membrane protein. Incubation was at 4 degrees for 1 h unless otherwise stated. The plates were blocked with 300 μl per well TBS-BSA for 1 h at room temperature. Subsequently, 100 μl of α-myc IgG (Sigma, 1:2000 in TBS-BSA) were added to each well and incubated for 2 h. DARPins were immobilized via their C-terminal myc-tag by applying 100 μl of crude extract diluted 1:5 in TBS-BSA. Binding of chemically or enzymatically biotinylated RND transporter (50 ng/μl in TBS-D-BSA, 100 μl/well) was detected using a Streptavidin alkaline phosphatase conjugate (Roche, 1:5000 in TBS-D-BSA, 100 μl/well) and the substrate di-sodium 4-nitrophenyl phosphate (Fluka). The absorbance at 405 nm was read using a Genion plate reader. In case of competition ELISA, the biotinylated RND-transporter was preincubated with a 20 fold excess of the competitor for 30 min at 4 degrees.

4.5.4 Stoichiometry

We determined the stoichiometry of AcrB-DARPin complexes by quantifying the molar concentration of each component by on-chip capillary electrophoresis using the Agilent Protein 80 or 230 Kit following the manufacturer's protocols (Agilent Technologies). The samples were prepared by gel filtration in buffer A/10 % glycerol to separate the complex from excess DARPin. The protein concentration of samples used for standard curves was determined by amino acid analysis. In order to compare the results from the two chips, we added 1/10 volume of MBP from a 1 mg/ml stock to each sample. A standard curve was prepared loading AcrB samples at different concentrations on the Agilent Protein 230 chip and we calculated the concentration of AcrB in the complex, which was typically around 3 μM. The same sample was analyzed on the Agilent Protein 80 chip, deducing the concentration of DARPin in the complex from a standard curve. Subsequently, the stoichiometry was calculated from the concentration ratio.

4.5.5 Surface plasmon resonance (Biacore, Proteon)

Surface plasmon resonance measurements to study the binding kinetics of DARPins to AcrB were carried out on a BIAcore T100 instrument at the Functional Genomics Center (FGCZ) of the University and ETH Zurich. Buffer A was used for immobilization as well as during the experiment. We immobilized 700 response units of chemically biotinylated AcrB in one flow cell of a SA chip (BIAcore) at a flow rate of 10 μl/min, leaving the other flow cell uncoated for referencing. DARPin analyte at various concentrations (0, 0.5, 1, 2, 4, 8, 16 and 32 nM or 0, 5, 10, 20, 40, 80, 160 and 320 nM) was injected with 180 s contact time and 600 s dissociation time at a flow rate of 60 μl/min. The dissociation time was sufficient to regenerate the chip. The data was analyzed using a kinetic fit with a 1:1 binding model after subtraction of the signal in the uncoated reference cell.

For the specificity analysis, surface plasmon resonance was analyzed using a ProteOn XPR36 protein interaction array system (BioRAD) at the FGCZ. On a NLC sensor chip (BioRAD), 1500 to 2000 RU of chemically biotinylated AcrB, AcrD, AcrF or MexB were immobilized via the interaction with neutravidin in ligand channels 1 to 4. The residual ligand channels were left uncoated. DARPins at 0, 3, 9, 27, 81 and 243 nM or 0, 15, 45, 135, 405 and 1215 nM were used for simultaneous analysis of the interaction in analyte channels 1 to 6. Binding kinetics were followed during 180 s contact time and 600 s dissociation time at a flow rate of 60 μl/min. The signal after double referencing using interspot regions and the blank analyte was analyzed qualitatively comparing the different RND transporters. For controls, DARPins specific for AcrF or

MexB were used. These had been generated by ribosome display selection from a N2C library using the protocol described previously [16].

4.5.6 Crystallization and structure determination

AcrB:DARPin complexes were crystallized by sitting drop vapour diffusion in 8-12 % PEG 4000, 50 mM ADA pH 6.5, 200 mM (NH4)2SO4 as described previously [16]. Data were collected on a PILATUS detector (Dectra) at the Swiss Light Source beamline XS06DA of the Paul Scherrer Institute (Villigen, Switzerland). XDS [42] was used for processing. The structures were solved by molecular replacement in PHASER [43] using the previously determined AcrB-DARPin complex as a search model (PDB ID 2J8S [16]) and refined in PHENIX [28].

4.5.7 Computational analysis of structures

The Computational Interface Alanine Scanning server [29] was used to identify residues crucial for DARPin binding to AcrB. The coordinate files for the structures of AcrB-DARPin complexes were submitted to the PISA server [44] at the European Bioinformatics Institute to calculate the buried surface area and analyze in detail the protein-protein interaction both in the DARPin-AcrB binding site and crystal contacts. Shape complementarity values were obtained using the program sc [30]. All figures of protein structures were prepared with Pymol [45].

4.5.8 Accession numbers

The atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession number 3NOC for the AcrB-DARPin#2 complex and 3NOG for AcrB-DARPin#3 complex.

Acknowledgements

We thank the staff of beamline PX of the Swiss Light Source, Villigen, Switzerland, for excellent technical assistance. This work was supported by the Swiss National Center of Competence in Research (NCCR) in Structural Biology.

Bibliography

[1] Andrew Ward, Christopher L Reyes, Jodie Yu, Christopher B Roth, and Geoffrey Chang. Flexibility in the ABC transporter MsbA: alternating access with a twist. *PNAS*, 104(48):19005–19010, 2007.

[2] Carola Hunte and Hartmut Michel. Crystallisation of membrane proteins mediated by antibody fragments. *Curr. Opin. Struct. Biol.*, 12:503–508, 2002.

[3] Y Zhou, J H Morais-Cabral, A Kaufman, and R MacKinnon. Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 å resolution. *Nature*, 414:43–48, 2001.

[4] Shohei Koide. Engineering of recombinant crystallization chaperones. *Curr. Opin. Struct. Biol.*, 19:449–457, 2009.

[5] H Kaspar Binz, Michael T Stumpp, Patrik Forrer, Patrick Amstutz, and Andreas Plückthun. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *JMB*, 332:489–503, 2003.

[6] H Kaspar Binz, Andreas Kohl, Andreas Plückthun, and Markus G Grütter. Crystal structure of a consensus-designed ankyrin repeat protein: implications for stability. *Proteins*, 65:280–284, 2006.

[7] J Hanes and A Plückthun. In vitro selection and evolution of functional proteins by using ribosome display. *PNAS*, 94:4937–4942, 1997.

[8] Daniel Steiner, Patrik Forrer, and Andreas Plückthun. Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *JMB*, 382:1211–1227, 2008.

[9] Patrick Amstutz, H Kaspar Binz, Petra Parizek, Michael T Stumpp, Andreas Kohl, Markus G Grütter, Patrik Forrer, and Andreas Plückthun. Intracellular kinase inhibitors

selected from combinatorial libraries of designed ankyrin repeat proteins. *JBC*, 280:24715–24722, 2005.

[10] H Kaspar Binz, Patrick Amstutz, Andreas Kohl, Michael T Stumpp, Christophe Briand, Patrik Forrer, Markus G Grütter, and Andreas Plückthun. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.*, 22:575–582, 2004.

[11] Thomas Huber, Daniel Steiner, Daniela Röthlisberger, and Andreas Plückthun. In vitro selection and characterization of DARPins and fab fragments for the co-crystallization of membrane proteins: The Na^{+}-citrate symporter CitS as an example. *J Struct Biol*, 159:206–221, 2007.

[12] Tiago M Bandeiras, Roman Christian Hillig, Pedro M Matias, Uwe Eberspaecher, Jörg Fanghänel, Mónica Thomaz, Sandra Miranda, Kerstin Crusius, Vera Pütter, Patrick Amstutz, Maya Gulotti-Georgieva, H Kaspar Binz, Caterina Holz, Arndt A P Schmitz, Christine Lang, Peter Donner, Ursula Egner, Maria A Carrondo, and Beate Müller-Tiemann. Structure of wild-type Plk-1 kinase domain in complex with a selective DARPin. *Acta Cryst D*, 64:339–353, 2008.

[13] Olivera Grubisha, Monika Kaminska, Stéphane Duquerroy, Elisabeth Fontan, Florence Cordier, Ahmed Haouz, Bertrand Raynal, Jeanne Chiaravalli, Muriel Delepierre, Alain Israël, Michel Véron, and Fabrice Agou. DARPin-assisted crystallography of the CC2-LZ domain of NEMO reveals a coupling between dimerization and ubiquitin binding. *JMB*, 395:89–104, 2010.

[14] Andreas Kohl, Patrick Amstutz, Petra Parizek, H Kaspar Binz, Christophe Briand, Guido Capitani, Patrik Forrer, Andreas Plückthun, and Markus G Grütter. Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure*, 13:1131–1141, 2005.

[15] Andreas Schweizer, Heidi Roschitzki-Voser, Patrick Amstutz, Christophe Briand, Maya Gulotti-Georgieva, Eva Prenosil, H Kaspar Binz, Guido Capitani, Antonio Baici, Andreas Plückthun, and Markus G Grütter. Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism. *Structure*, 15:625–636, 2007.

[16] Gaby Sennhauser, Patrick Amstutz, Christophe Briand, Otso Storchenegger, and Markus G Grütter. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biology*, 5:e7, 2007.

[17] David Veesler, Birgit Dreier, Stéphanie Blangy, Julie Lichière, Denise Tremblay, Sylvain Moineau, Silvia Spinelli, Mariella Tegoni, Andreas Plückthun, Valérie Campanacci, and Christian Cambillau. Crystal structure and function of a DARPin neutralizing inhibitor of lactococcal phage TP901-1: comparison of DARPin and camelid VHH binding mode. *JBC*, 284:30718–30726, 2009.

[18] Gaby Sennhauser and Markus G Grütter. Chaperone-assisted crystallography with DARPins. *Structure*, 16:1443–1453, 2008.

[19] T T Tseng, K S Gratwick, J Kollman, D Park, D H Nies, A Goffeau, and M H Saier. The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.*, 1(1):107–125, 1999.

[20] Hiroshi Nikaido and Yumiko Takatsuka. Mechanisms of RND multidrug efflux pumps. *Biochimica Et Biophysica Acta*, 1794:769–781, 2009.

[21] Satoshi Murakami, Ryosuke Nakashima, Eiki Yamashita, and Akihito Yamaguchi. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, 419:587–593, 2002.

[22] Klaas M Pos, André Schiefner, Markus A Seeger, and Kay Diederichs. Crystallographic analysis of AcrB. *FEBS Letters*, 564(3):333–339, 2004.

[23] Satoshi Murakami, Ryosuke Nakashima, Eiki Yamashita, Takashi Matsumoto, and Akihito Yamaguchi. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature*, 443:173–179, 2006.

[24] Markus A Seeger, André Schiefner, Thomas Eicher, François Verrey, Kay Diederichs, and Klaas M Pos. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*, 313:1295–1298, 2006.

[25] Markus A Seeger, Christoph von Ballmoos, Thomas Eicher, Lorenz Brandstätter, François Verrey, Kay Diederichs, and Klaas M Pos. Engineered disulfide bonds support the functional rotation mechanism of multidrug efflux pump AcrB. *Nat Struct Mol Biol*, 15:199–205, 2008.

[26] Michaela A Kramer, Svava K Wetzel, Andreas Plückthun, Peer R E Mittl, and Markus G Grütter. Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned c-capping module. *JMB*, 404(3):381–391, 2010.

[27] E Beitz. TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics*, 16:135–139, 2000.

[28] Paul D Adams, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, Li-Wei Hung, Gary J Kapral, Ralf W Grosse-Kunstleve, Airlie J McCoy, Nigel W Moriarty, Robert Oeffner, Randy J Read, David C Richardson, Jane S Richardson, Thomas C Terwilliger, and Peter H Zwart. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D*, 66:213–221, 2010.

[29] Tanja Kortemme, David E Kim, and David Baker. Computational alanine scanning of protein-protein interfaces. *Science's STKE: Signal Transduction Knowledge Environment*, 2004:pl2, 2004.

[30] M C Lawrence and P M Colman. Shape complementarity at protein/protein interfaces. *JMB*, 234:946–950, 1993.

[31] Anthony A Kossiakoff and Shohei Koide. Understanding mechanisms governing protein-protein interactions from synthetic binding interfaces. *Curr. Opin. Struct. Biol.*, 18:499–506, 2008.

[32] Gábor Pál, Jean-Louis K Kouadio, Dean R Artis, Anthony A Kossiakoff, and Sachdev S Sidhu. Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *JBC*, 281:22378–22385, 2006.

[33] I. Saira Mian, Arthur R. Bradwell, and Arthur J. Olson. Structure, function and properties of antibody binding sites. *JMB*, 217:133–151, 1991.

[34] Frederic A Fellouse, Christian Wiesmann, and Sachdev S Sidhu. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *PNAS*, 101:12467–12472, 2004.

[35] Frederic A Fellouse, Bing Li, Deanne M Compaan, Andrew A Peden, Sarah G Hymowitz, and Sachdev S Sidhu. Molecular recognition by a binary code. *JMB*, 348:1153–1162, 2005.

[36] Akiko Koide, Ryan N Gilbreth, Kaori Esaki, Valentina Tereshko, and Shohei Koide. High-affinity single-domain binding proteins with a binary-code interface. *PNAS*, 104:6632–6637, 2007.

[37] Martyn F Symmons, Evert Bokma, Eva Koronakis, Colin Hughes, and Vassilis Koronakis. The assembled structure of a complete tripartite bacterial multidrug efflux pump. *PNAS*, 106:7173–7178, 2009.

[38] Klaas M Pos. Trinity revealed: Stoichiometric complex assembly of a bacterial multidrug efflux pump. *PNAS*, 106:6893–6894, 2009.

[39] Zygmunt S Derewenda and Peter G Vekilov. Entropy and surface engineering in protein crystallization. *Acta Cryst D*, 62:116–124, 2006.

[40] Zygmunt S. Derewenda. Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Cryst D*, 66(5):604–615, 2010.

[41] M G Cull and P J Schatz. Biotinylation of proteins in vivo and in vitro using small peptide tags. *Method Enzymol*, 326:430–440, 2000.

[42] W. Kabsch. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.*, 26:795–800, 1993.

[43] Airlie J McCoy, Ralf W Grosse-Kunstleve, Paul D Adams, Martyn D Winn, Laurent C Storoni, and Randy J Read. Phaser crystallographic software. *J. Appl. Crystallogr.*, 40:658–674, 2007.

[44] Evgeny Krissinel and Kim Henrick. Inference of macromolecular assemblies from crystalline state. *JMB*, 372:774–797, 2007.

[45] Warren L DeLano. *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA, 2002.

Appendix A

Materials

A.1 Cell lines and bacterial strains

Sf9 from *Spodoptera frugiperda*	Invitrogen
Escherichia coli DH10Bac$^{\text{TM}}$	F mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara leu) 7697 galU galK λ^- rpsL nupG/ pMON14272/pMON7124 Invitrogen
Escherichia coli XL1blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac FŠ proAB lacIqZΔM15 Tn10 Tet$^{\text{r}}$ Stratagene
Escherichia coli BL21(DE3)	F dcm ompT hsdS(r$_{\text{B}}$ m$_{\text{B}}$) gal λ(DE3) [1]
Escherichia coli C43(DE3)	F dcm ompT hsdS$_{\text{B}}$(r$_{\text{B}}$ m$_{\text{B}}$) gal lon λ(DE3) [2]

A.2 Oligonucleotides and plasmids

Oligonucleotides for the generation and analysis of Ntcp vectors and bacmids

BamHI-Ntcp	5′-CCA GGA TCC GAT GGA GGT GCA CAA CG-3′
Ntcp-HindIII	5′-CAC AAG CTT ATT ATC TAG AAT TTG CCA TCT GAC CAG-3′
Xba-His-for	5′-CTA GAC ACC ACC ACC ACC ACC ACT AAT A-3′
Xba-His-rev	5′-AGC TTA TTA GTG GTG GTG GTG GTG GTG T-3′
Strep-for	5′-CTA GAT GGT CGC ACC CGC AAT TTG AAA AAT

	AAT A-3′
Strep-rev	5′-AGC TTA TTA TTT TTC AAA TTG CGG GTG CGA CCA T-3′
FastBacf	5′-TAC TGT TTT CGT AAC AGT TTT G-3′
FastBacr	5′-CAT TTT ATG TTT CAG GTT CAG G-3′
M13-for	5′-GTT TTC CCA GTC ACG AC-3′
M13-rev	5′-CAG GAA ACA GCT ATG AC-3′

Oligonucleotides for the construction of expression vectors for extracellular domain of Oatp1a4

loop1a4-for1	5′-CCT CCA TGG AAT TCG TGG CAG GCT TAA C-3′
loop1a4-for2	5′-CCT CCA TGG AAT TCG CTG ACT GTA ACA C-3′
loop1a4-rev1	5′-CCT CTA GAA GCT TGG CAC AGT CAG GG-3′
loop1a4-rev2	5′-CCT CTA GAA GCT TAA TGC AGC TGC AAT TC-3′
loop1a4-rev1-stop	5′-GCA AGC TTT TAT TAC CCG GGG GCA CAG TCA GG-3′
loop1a4-rev2-stop	5′-GCA AGC TTT TAT TAC CCG GGA ATG CAG CTG CA-3′

Oligonucleotides for the modification and amplification of DARPins

EWT5	5′-TTC CTC CAT GGG TAT GAG AGG ATC GCA TCA CCA TCA CCA TCA CGG ATC CGA CCT GGG-3′
WTC4	5′-TTT GGG AAG CTT TTG CAG GAT TTC AGC-3′
MTS46	5′-GCA TAA GCT TCA TTA GTT CAG TTT CTG CAG GAT TTC AGC CAG-3′

Plasmids

pSPORT-Ntcp	Stieger lab, University Hospital Zurich
pSPORT-Oatp1a4	Stieger lab, University Hospital Zurich
pET28-AcrB-His	Gaby Sennhauser, University of Zurich [3]
pQE30	Qiagen
pFastBac	Invitrogen
pET20b(+)	Novagen
pMAL-c2E	New England Biolabs

pMAL-pm1	Peer Mittl, University of Zurich
pGEX-6P1	Amersham Biosciences
pETM80	Annett Dümmler, EMBL Heidelberg [4]
pETM82	Annett Dümmler, EMBL Heidelberg [4]
pETNus-1a	Sattler lab, EMBL Heidelberg
pETTrx-1a	Sattler lab, EMBL Heidelberg
pETGB1-1a	Sattler lab, EMBL Heidelberg
pETMBP-1a	Sattler lab, EMBL Heidelberg

A.3 Enzymes and antibodies

Enzymes for DNA modification

RNAse A	Roche
BamHI	New England Biolabs
HindIII	New England Biolabs
T4 DNA ligase	New England Biolabs
Taq polymerase	New England Biolabs

Enzymes used during protein purification

DNAseI	Roche
Tobacco Etch Virus (TEV) protease	produced and purified in-house

Reagents for immunoassays

Protein A	Sigma
Mouse α-myc IgG	Sigma
K4 rabbit α-Ntcp	polyclonal antibody provided by Stieger lab, University Hospital Zurich
Goat α-rabbit IgG horse radish peroxidase (HRP) conjugate	Sigma
Streptavidin alkaline phosphatase (AP) conjugate	Roche

Disodium 4-nitrophenyl phosphate (pNPP)	Fluka
ECl Kit	GE Healthcare

A.4 Media and solutions for cloning and protein expression

Media and stock solutions for bacterial expression

LB	10 g/l bacto tryptone 5 g/l yeast extract 10 g/l NaCl
2xYT	16 g/l bacto tryptone 10 g/l yeast extract 5 g/l NaCl
Ampicillin	100 mg/ml stock
Kanamycin	50 mg/ml stock
Tetracyclin	15 mg/ml stock
Gentamycin	7 mg/ml stock
X-gal	100 mg/ml stock
IPTG	1 M stock

Insect cell culture media, solutions and reagents

Sf-900 II SFM	Gibco
Grace's Medium	Gibco
Fetal Calf Serum (FCS)	Gibco
Penicillin/Streptomycin	Gibco
Cellfectin	Invitrogen
Trypan blue	Invitrogen
Thiazolyl Blue Tetrazolium Blue (MTT)	Sigma

A.5 Buffers and solutions

Buffers for the isolation of bacmid DNA

Solution I (resuspension)	15 mM Tris/HCl pH 8.0 10 mM EDTA 100 μM RNAse A filtered 0.2 μm
Solution II (lysis)	0.2 N NaOH 1 % (w/v) SDS filtered 0.2 μm
Solution III (neutralization)	3 M potassium acetate, pH 5.5 autoclaved

Buffers for the Ntcp transport assay in Sf9 cells

Sodium buffer	122 mM NaCl 10 mM MES pH 6.5 3 mM $CaCl_2$ 5 mM $MgCl_2$ 5 mM D-glucose
Choline buffer	122 mM choline chloride 10 mM MES pH 6.5 3 mM $CaCl_2$ 5 mM $MgCl_2$ 5 mM D-glucose
Hot solution	10 μM [^{3}H]-taurocholate (2.5 μCi/ml in sodium or choline buffer

Buffers for membrane preparation from Sf9 cells expressing Ntcp

TMEP	50 mM Tris/HCl pH 7.0
	50 mM mannitol
	2 mM EGTA
	Complete (Roche)
SMS	20 mM HEPES/Tris pH 7.5
	250 mM sucrose
	0.2 mM $\mathrm{CaCl_2}$

Buffers for the purification of the extracellular domain of Oatp1a4

Amylose lysis buffer	50 mM Tris/HCl pH 8.0
	150 mM NaCl
	1 mM DTT
	0.1 mg/ml DNAse I
Amylose column running buffer	50 mM Tris/HCl pH 8.0
	150 mM NaCl
Amylose column elution buffer	50 mM Tris/HCl pH 8.0
	150 mM NaCl
	10 mM D-maltose
Ni-NTA ysis buffer	50 mM Tris/HCl pH 8.0
	150 mM NaCl
	3 mM GSH
	3 mM GSSG
	10 mM imidazole
	0.1 mg/ml DNAse I
Ni-NTA loading buffer	50 mM Tris/HCl pH 8.0
	150 mM NaCl

	10 mM imidazole
Ni-NTA wash buffer	Ni-NTA buffer with 20 mM imidazole
Ni-NTA elution buffer	Ni-NTA buffer with 250 mM imidazole
TEV cleavage buffer	50 mM Tris/HCl pH 8.0 150 mM NaCl 3 mM GSH 0.3 mM GSSG
or	50 mM Tris/HCl pH 8.0 150 mM NaCl 1 mM DTT
Size exclusion buffer	50 mM Tris/HCl pH 8.0 150 mM NaCl

Buffers for the purification of AcrB

Master buffer (Buffer A)	20 mM Tris/HCl pH 7.5 150 mM NaCl
Lysis buffer (LB)	20 mM Tris/HCl pH 7.5 150 mM NaCl 0.1 mg/ml DNAseI Complete
Solubilization buffer (SB)	20 mM Tris/HCl pH 7.5 150 mM NaCl 10 % (v/v) glycerol 10 mM imidazole 1 % DDM Complete (EDTA-free)

Binding buffer (BB)	20 mM Tris/HCl pH 7.5 150 mM NaCl 10 % (v/v) glycerol 10 mM imidazole 0.03 % DDM
Wash buffer (WB)	BB with 50 mM imidazole
Elution buffer (EB)	BB with 200 mM imidazole
Gel filtration buffer (GF)	20 mM Tris/HCl pH 7.5 150 mM NaCl 0.03 % DDM
HEPES buffer	10 mM HEPES/NaOH pH 7.5 50 mM NaCl 0.03 % DDM

Buffers for the purification of DARPins

Lysis buffer	50 mM Tris/HCl pH 7.5 400 mM NaCl 0.1 mg/ml DNAseI
Binding buffer	50 mM Tris/HCl pH 7.5 400 mM NaCl 20 mM imidazole 10 % (w/v) glycerol
Elution buffer	50 mM Tris/HCl pH 7.5 400 mM NaCl 250 mM imidazole 10 % (w/v) glycerol

Buffers for SDS PAGE and Western Blot

Electrophoresis buffer (5x)	125 mM Tris 1 M glycine 5 g/l (w/v) SDS
high TBS	50 mM Tris/HCl pH 7.0 500 mM NaCl
high TBS-T	TBS with 0.2 % Tween20

Bibliography

[1] F W Studier and B A Moffatt. Use of bacteriophage t7 RNA polymerase to direct selective high-level expression of cloned genes. *JMB*, 189:113–130, 1986.

[2] B Miroux and J E Walker. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *JMB*, 260(3):289–298, 1996.

[3] Gaby Sennhauser, Patrick Amstutz, Christophe Briand, Otso Storchenegger, and Markus G Grütter. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biology*, 5:e7, 2007.

[4] Annett Dümmler, Ann-Marie Lawrence, and Ario de Marco. Simplified screening for the detection of soluble fusion constructs expressed in e. coli using a modular set of vectors. *Microbial Cell Factories*, 4:34, 2005.

Appendix B

Crystallization of other AcrB-DARPin complexes

Further crystallization trials were carried out with AcrB-DARPin complexes. Figure B.1 and Table B.1 summarize hits observed during the screening process. All crystals were tested at the Swiss Light Source (Paul-Scherrer Institute, Villigen, Switzerland). Some crystals diffracted to moderate resolution. Where the data was sufficient to determine the unit cell dimensions, the processing statistics are listed in Table (insert Table). For the two cases where the structures could be solved by molecular replacement with AcrB at low resolution, no density for a DARPin binder was found.

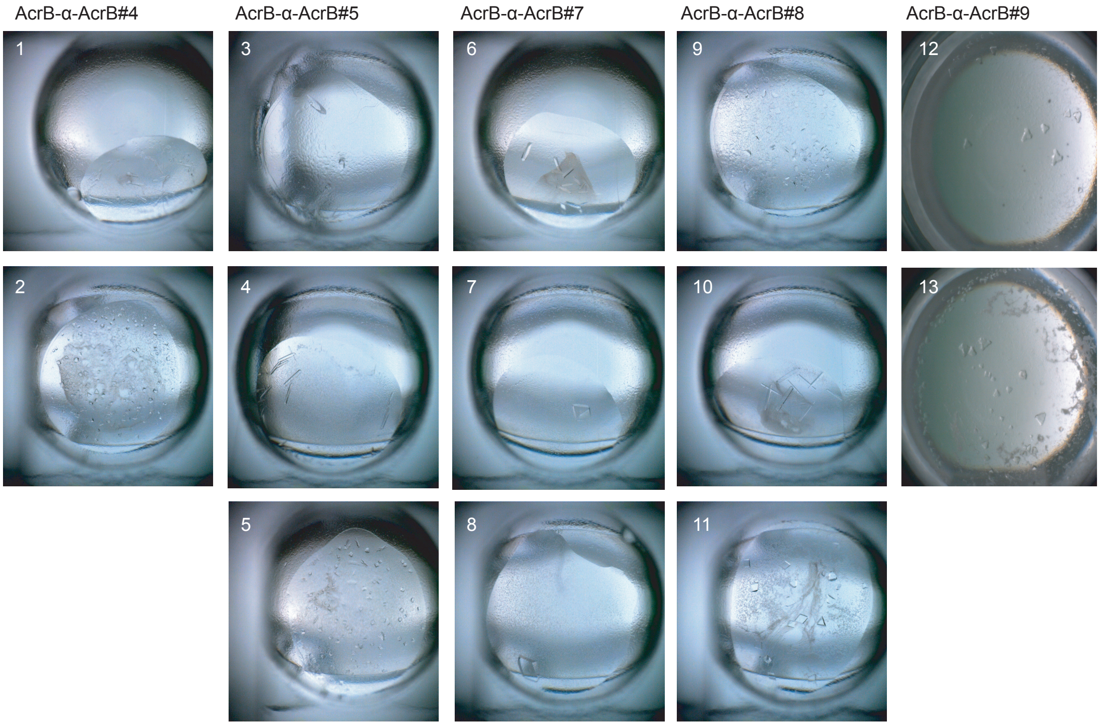


Figure B.1: Crystallization of AcrB in complex with DARPin#4, #5, #7, #8 and #9. The conditions are listed in Table B.1.

Table B.1: Crystallization of AcrB-DARPin complexes. This table summarizes the crystallization conditions and diffraction properties of the AcrB-DARPin crystals shown in Figure B.1.

Image	Complex	Condition	Comments
1	AcrB-DARPin#4	15 % PEG400, 50 mM glycine pH 9.4, 150 mM NaCl	
2	AcrB-DARPin#4	10 % PEG4000, 50 mM Na cacodylate pH 5.5, 1 M NaCl	
3	AcrB-DARPin#5	5 % PEG4000, 50 mM Na cacodylate pH 5.5, 150 mM magnesium acetate	diffracted to 20 Å

4	AcrB-DARPin#5	5 % PEG4000, 50 mM Na cacodylate pH 5.5, 175 mM magnesium acetate	
5	AcrB-DARPin#5	15 % PEG400, 50 mM ADA pH 6.5, 50 mM magnesium acetate	
6	AcrB-DARPin#7	15 % PEG400, 50 mM Na acetate pH 4.5, 50 mM magnesium acetate	diffracted to 7 Å
7	AcrB-DARPin#7	5 % PEG4000, 50 mM ADA pH 6.5, 200 mM ammonium sulfate	diffracted to 5 Å
8	AcrB-DARPin#7	10 % PEG4000, 50 mM ADA pH 6.5, 200 mM ammonium sulfate	
9	AcrB-DARPin#8	5 % PEG4000, 50 mM ADA pH 6.5, 1 M NaCl	
10	AcrB-DARPin#8	5 % PEG4000, 50 mM Na acetate pH 4.5, 150 mM NaCl	20 frames
11	AcrB-DARPin#8	12 % PEG4000, 100 mM MES pH 6.5, 100 mM NaCl	
12	AcrB-DARPin#9	30 % PEG400, 100 mM ADA pH 6.5, 300 mM lithium sulfate, 2 mM dequalinium	diffracted to 8.5 Å

Appendix C

Curriculum vitae

Personal Information

Name	Nicole Monroe
Maiden name	Bischoff
Date of birth	March 9, 1981
Place of birth	Mannheim, Germany

Education

11/2005 to present	**University of Zurich, Switzerland** PhD program in Molecular Life Sciences
10/2004 to 07/2005	**University of Oregon, USA** Master thesis under Prof. Brian Matthews Title: A molecular switch in T4 lysozyme
09/2003 to 09/2005	**Technical University Munich, Germany** Graduate program in Molecular Biotechnology Degree: Master of Science Major: Protein structure, function and engineering Minors: Biochemical engineering, business economics

09/2000 to 09/2003 **Technical University Munich, Germany**
Undergraduate program in Molecular Biotechnology
Degree: Bachelor of Science

10/2004 to 07/2005 **Technical University Munich, Germany**
Bachelor thesis under Prof. Arne Skerra
Title: Implementation eines neuen Bindungsmechanismus im Retinolbindungsprotein

2000 **Illertalgymnasium Vöhringen, Germany**
Degree: Allgemeine Hochschulreife

Publications

Monroe N, Sennhauser G, Seeger MA, Briand C, Grütter MG (2011): Designed ankyrin repeat protein binders for the crystallization of AcrB: Plasticity of the dominant interface. *J. Struct. Biol.* 174(2):269-81

Yousef MA, Bischoff N, Dyer CM, Baase WA, Matthews BW (2006): Guanidinium derivatives bind preferentially and trigger long-distance conformational changes. *Protein Sci.* 14(4):853-61

Workshops

09/2008 PEPC6: Protein expression, purification and crystallization
EMBL Hamburg, Germany

05/2008 CCP4 school: From data processing to structure refinement and beyond
Argonne National Laboratory, IL, USA

08/2007 Diffraction and data collection
BESSY, Berlin, Germany

Conferences

09/2009	Rhine-Knee Regio Meeting, Leissingen, Switzerland Oral presentation: DARPin binders to AcrB: Plasticity of the interface
09/2009	International NCCR Symposium on New Trends in Structural Biology Zurich, Switzerland Poster presentation
03/2009	8th Day of Clinical Research University Hospital Zurich, Switzerland Poster presentation
09/2008	International NCCR Symposium on New Trends in Structural Biology Zurich, Switzerland Poster presentation
07/2007	Biomedical Transporters: Membrane Transporters in Disease and Drug Development Bern, Switzerland
09/2007	Annual retreat of the Molecular Life Science PhD program Fiescheralp, Switzerland Poster presentation
10/2006	Annual retreat of the Molecular Life Science PhD program Chandolin, Switzerland Poster presentation